GenCore version 5.1.9
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 15, 2006, 18:04:58 ; Search time 547 Seconds
                                        (without alignments)
                                        7647.805 Million cell updates/sec

Title: US-10-722-939-1-T39977_39901-40500
Perfect score: 600
Sequence: 1 tacaatagaccctgcttctt..........aataagtgacagagctgtga 600

Scoring table: IDENTITY_NUC
               Gapop 10.0 , Gapext 1.0

Searched: 5244920 seqs, 3486124231 residues

Total number of hits satisfying chosen parameters: 10489840

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 100 summaries

Database :   N_Geneseq_8:*
               1: geneseqn1980s:*
               2: geneseqn1990s:*
               3: geneseqn2000s:*
               4: geneseqn2001as:*
               5: geneseqn2001bs:*
               6: geneseqn2002as:*
               7: geneseqn2002bs:*
               8: geneseqn2003as:*
               9: geneseqn2003bs:*
               10: geneseqn2003cs:*
               11: geneseqn2003ds:*
               12: geneseqn2004as:*
               13: geneseqn2004bs:*
               14: geneseqn2005s:*
               15: geneseqn2006s:*

         Pred. No. is the number of results predicted by chance to have a
         score greater than or equal to the score of the result being printed,
         and is derived by analysis of the total score distribution.

                                SUMMARIES
                       %
Result   Query
   No.   Score Match Length DB   ID   Description
       ----
       1   599.6 100.0 76600 12   ADO79402   Ado79402 DLG1 regi
c   2   98   16.3   2778 15   AEE84582   Aee84582 Discs lar

c	3	98	16.3	2980	14	AED21168	Aed21168 Human Dlg
c	4	98	16.3	2980	15	AEF75023	Aef75023 Human pol
c	5	98	16.3	3046	9	ACC85071	Acc85071 Human MBC
c	6	98	16.3	3046	10	ADK11476	Adk11476 Human dis
c	7	98	16.3	3046	13	ADR83538	Adr83538 Human dis
c	8	98	16.3	3046	13	ADU05806	Adu05806 Novel bro
c	9	97.6	16.3	3046	12	ADO79406	Ado79406 DLG1 cDNA
c	10	96.4	16.1	1251	5	AAC90758	Aac90758 Mammalian
c	11	96.4	16.1	4108	9	AAD57342	Aad57342 Human kin
c	12	96	16.0	1152	4	AAH34453	Aah34453 Human col
c	13	82.6	13.8	419	8	ABX46069	Abx46069 Bovine ES
c	14	76.8	12.8	8056	8	ABZ10246	Abz10246 Haematopo
c	15	76.2	12.7	3256	13	ADV40930	Adv40930 Rat cardi
c	16	76.2	12.7	3256	14	AEA47948	Aea47948 Rat synap
c	17	75.2	12.5	8056	8	ABZ10100	Abz10100 Haematopo
c	18	74.8	12.5	516	8	ABX40620	Abx40620 Bovine ES
	19	73.6	12.3	612	14	ACL64894	Acl64894 M. xanthu
c	20	73.2	12.2	6145	6	ABL32972	Abl32972 Human imm
c	21	72.8	12.1	6292	4	AAS46735	Aas46735 Tumour su
c	22	72.6	12.1	14615	4	AAS46704	Aas46704 Tumour su
c	23 ′	71.4	11.9	3150	14	AED21170	Aed21170 Mouse Dlg
c	24	71.2	11.9	5845	6	ABL33662	Abl33662 Human imm
c	25	71.2	11.9	6161	6	ABL32411	Abl32411 Human imm
c	26	71	11.8	363	5	ADI71847	Adi71847 Human ova
c	27	71	11.8	363	5	ADL36998	Adl36998 Human ova
c	28	70.4	11.7	547	13	ACN62109	Acn62109 Cotton gy
c	29	70.4	11.7	110000	5	AAI61373_0	Aai61373 Soybean 3
c	30	70.2	11.7	16766	6	ABL34157	Abl34157 Human imm
c	31	70	11.7	3683	8	ABZ10199	Abz10199 Haematopo
c	32	70	11.7	12237	6	ABL34358	Abl34358 Human imm
c	33	69.8	11.6	83391	6	ABQ67093	Abq67093 Human ang
c	34	69.6	11.6	7442	4	AAS46686	Aas46686 Tumour su
c	35	69.2	11.5	408	8	ABX36056	Abx36056 Bovine ES
c	36	69.2	11.5	5413	6	ABL32564	Abl32564 Human imm
c	37	69	11.5	5750	4	AAS46708	Aas46708 Tumour su
c	38	69	11.5	5750	6	ABL34008	Abl34008 Human imm
c	39	69	11.5	6154	6	ABK31200	Abk31200 Signal tr
c	40	69	11.5	6154	6	ABL70167	Abl70167 Chemicall
c	41	69	11.5	6154	6	AAS61112	Aas61112 Human gen
c	42	69	11.5	8900	13	ADS89685	Ads89685 Oligonucl
c	43	68.8	11.5	5572	6	ABL33425	Abl33425 Human imm
c	44	68.8	11.5	5572	6	ABL54360	Abl54360 Chemicall
c	45	68.8	11.5	7657	4	AAS45477	Aas45477 Chemicall
c	46	68.8	11.5	7657	6	ABL34022	Abl34022 Human imm
	47	68.6	11.4	9789	2	AAT41852	Aat41852 cDNA enco
c	48	68.6	11.4	40324	6	ABQ67149	Abq67149 Human ang
	49	68.6	11.4	48933	11	ACN44406	Acn44406 Human gen
	50	68.2	11.4	8056	8	ABZ10100	Abz10100 Haematopo
c	51	67.8	11.3	393	8	ABX39417	Abx39417 Bovine ES
c	52	67.8	11.3	12405	4	AAS45330	Aas45330 Chemicall
c	53	67.8	11.3	12405	6	ABK28169	Abk28169 DNA trans
c	54	67.8	11.3	12405	6	AAS61143	Aas61143 Human gen
c	55	67.6	11.3	314	8	ABX47247	Abx47247 Bovine ES
c	56	67.6	11.3	6775	6	ABQ67159	Abq67159 Human ang
	57	67.6	11.3	8056	8	ABZ10246	Abz10246 Haematopo
c	58	67.6	11.3	8900	13	ADS89411	Ads89411 Oligonucl
c	59	67.6	11.3	26997	4	AAS46747	Aas46747 Tumour su

```
c  60    67.4   11.2   10891   6   ABL32464        Abl32464 Human imm
c  61    67.4   11.2   11097  10   ADF50896        Adf50896 Chemicall
c  62    67.4   11.2   11745   6   ABK28331        Abk28331 DNA trans
c  63    67.2   11.2    5406   6   ABL33098        Abl33098 Human imm
c  64    67.2   11.2    7321   6   ABK31425        Abk31425 Signal tr
c  65    67.2   11.2    7321   6   ABL70386        Abl70386 Chemicall
c  66    67.2   11.2    7321   6   AAS61338        Aas61338 Human gen
c  67    67.2   11.2   15548   6   ABL34155        Abl34155 Human imm
c  68    67     11.2     495  13   ACN47898         Acn47898 Cotton pr
c  69    67     11.2    6552   4   AAS46678        Aas46678 Tumour su
c  70    67     11.2    6767   4   AAS46608        Aas46608 Tumour su
c  71    67     11.2   14987   6   ABL32631        Abl32631 Human imm
   72    67     11.2   94330  11   ACN44662         Acn44662 Human gen
c  73    66.8   11.1     262   7   ADS31066        Ads31066 Human gen
c  74    66.8   11.1     262   7   ADY36454        Ady36454 HIRA geno
c  75    66.8   11.1     446   8   ABX38235        Abx38235 Bovine ES
c  76    66.8   11.1     560  13   ACN47472         Acn47472 Cotton pr
c  77    66.8   11.1    6171   6   ABL32788        Abl32788 Human imm
c  78    66.8   11.1  113515   6   ABL34175        Abl34175 Human imm
c  79    66.6   11.1   14006   6   ABL33958        Abl33958 Human imm
c  80    66.2   11.0    8186   6   ABN80090        Abn80090 Human che
c  81    66     11.0     375   5   ABV44911        Abv44911 Human pro
c  82    66     11.0    3780  13   ADS89381         Ads89381 Oligonucl
c  83    66     11.0    3780  13   ADS89655         Ads89655 Oligonucl
c  84    65.8   11.0     349   5   ADL44149        Adl44149 Human ova
c  85    65.8   11.0     578  13   ACN52669         Acn52669 Cotton an
c  86    65.8   11.0    2739   6   ABQ67176        Abq67176 Human ang
c  87    65.8   11.0    3315   8   ACF62784        Acf62784 Colon can
c  88    65.8   11.0    3315   8   ACF62806        Acf62806 Colon can
c  89    65.8   11.0    3814   8   ABZ10010        Abz10010 Haematopo
c  90    65.8   11.0    3814   8   ABZ10156        Abz10156 Haematopo
c  91    65.8   11.0    3814  13   ADS89370         Ads89370 Oligonucl
c  92    65.8   11.0    3814  13   ADS89616         Ads89616 Oligonucl
c  93    65.8   11.0    3814  13   ADS89644         Ads89644 Oligonucl
c  94    65.8   11.0    3814  13   ADS89342         Ads89342 Oligonucl
c  95    65.8   11.0    5476   4   AAS45499        Aas45499 Chemicall
c  96    65.8   11.0    5476   6   ABL34114        Abl34114 Human imm
c  97    65.8   11.0    5476   6   ABL49381        Abl49381 Human pol
c  98    65.8   11.0    5888   6   ABL34456        Abl34456 Human met
c  99    65.8   11.0    5888   7   ADS99717        Ads99717 Bisulphit
c 100    65.8   11.0    6109   6   ABL32326        Abl32326 Human imm
```

                              ALIGNMENTS

RESULT 1
ADO79402
ID   ADO79402 standard; DNA; 76600 BP.
XX
AC   ADO79402;
XX
DT   26-AUG-2004   (first entry)
XX
DE   DLG1 region, SEQ ID 1.
XX
KW   Cytostatic; Gene therapy; breast cancer; human; DLG1; KIAA0783; DPF3;

```
KW    CENPC1; gene; ds; SNP; single nucleotide polymorphism;
KW    synapse-associated protein 97; hdlg; SAP97; chromosome 3q29.
XX
OS    Homo sapiens.
XX
FH    Key             Location/Qualifiers
FT    variation       133
FT                    /*tag=  a
FT                    /standard_name= "Single nucleotide polymorphism"
FT                    /note= "This SNP is described as a T/C SNP"
FT    variation       7938
FT                    /*tag=  b
FT                    /standard_name= "Single nucleotide polymorphism"
FT                    /note= "This SNP is described as a T/C SNP"
FT    variation       8873
FT                    /*tag=  c
FT                    /standard_name= "Single nucleotide polymorphism"
FT                    /note= "This SNP is described as a G/A SNP"
FT    variation       13221
FT                    /*tag=  d
FT                    /standard_name= "Single nucleotide polymorphism"
FT                    /note= "This SNP is described as a T/C SNP"
FT    variation       17288
FT                    /*tag=  e
FT                    /standard_name= "Single nucleotide polymorphism"
FT                    /note= "This SNP is described as a T/C SNP"
FT    variation       25732
FT                    /*tag=  f
FT                    /standard_name= "Single nucleotide polymorphism"
FT                    /note= "This SNP is described as a G/A SNP"
FT    variation       26923
FT                    /*tag=  g
FT                    /standard_name= "Single nucleotide polymorphism"
FT                    /note= "This SNP is described as a C/G SNP"
FT    variation       39977
FT                    /*tag=  h
FT                    /standard_name= "Single nucleotide polymorphism"
FT                    /note= "This SNP is described as a T/C SNP"
FT    variation       41284
FT                    /*tag=  i
FT                    /standard_name= "Single nucleotide polymorphism"
FT                    /note= "This SNP is described as a T/A SNP"
FT    variation       41410
FT                    /*tag=  j
FT                    /standard_name= "Single nucleotide polymorphism"
FT                    /note= "This SNP is described as a A/C SNP"
FT    variation       41477
FT                    /*tag=  k
FT                    /standard_name= "Single nucleotide polymorphism"
FT                    /note= "This SNP is described as a T/C SNP"
FT    variation       41514
FT                    /*tag=  l
FT                    /standard_name= "Single nucleotide polymorphism"
FT                    /note= "This SNP is described as a G/A SNP"
FT    variation       42606
FT                    /*tag=  m
FT                    /standard_name= "Single nucleotide polymorphism"
```

```
FT                      /note= "This SNP is described as a G/A SNP"
FT     variation        42742
FT                      /*tag=  n
FT                      /standard_name= "Single nucleotide polymorphism"
FT                      /note= "This SNP is described as a A/C SNP"
FT     variation        59515
FT                      /*tag=  o
FT                      /standard_name= "Single nucleotide polymorphism"
FT                      /note= "This SNP is described as a G/A SNP"
FT     variation        59808
FT                      /*tag=  p
FT                      /standard_name= "Single nucleotide polymorphism"
FT                      /note= "This SNP is described as a T/C SNP"
FT     variation        60265
FT                      /*tag=  q
FT                      /standard_name= "Single nucleotide polymorphism"
FT                      /note= "This SNP is described as a C/G SNP"
FT     variation        67152
FT                      /*tag=  r
FT                      /standard_name= "Single nucleotide polymorphism"
FT                      /note= "This SNP is described as a T/C SNP"
FT     variation        68332
FT                      /*tag=  s
FT                      /standard_name= "Single nucleotide polymorphism"
FT                      /note= "This SNP is described as a T/C SNP"
FT     variation        71128
FT                      /*tag=  t
FT                      /standard_name= "Single nucleotide polymorphism"
FT                      /note= "This SNP is described as a T/C SNP"
FT     variation        76427
FT                      /*tag=  u
FT                      /standard_name= "Single nucleotide polymorphism"
FT                      /note= "This SNP is described as a G/A SNP"
XX
PN     WO2004047514-A2.
XX
PD     10-JUN-2004.
XX
PF     25-NOV-2003; 2003WO-US037943.
XX
PR     25-NOV-2002; 2002US-0429136P.
PR     24-JUL-2003; 2003US-0490234P.
XX
PA     (SEQU-) SEQUENOM INC.
XX
PI     Roth RB,  Nelson MR,  Braun A,  Kammerer SM,  Reneland R;
XX
DR     WPI; 2004-441037/41.
XX
PT     Identifying a subject at risk of breast cancer by detecting the presence
PT     of polymorphic variations in the DLG1, KIAA0783, DPF3 or CENPC1 regions
PT     which are associated with breast cancer in a nucleic acid sample from a
PT     subject.
XX
PS     Claim 24; Fig 1; 227pp; English.
XX
CC     The present invention relates to a method for identifying a subject at
```

CC   risk of breast cancer. The method comprising detecting the presence or
CC   absence of one or more polymorphic variations associated with breast
CC   cancer in a nucleic acid sample from a subject. The nucleic acid sample
CC   comprises the DLG1 region (ADO79402), KIAA0783 region (ADO79403), DPF3
CC   region (ADO79404) or CENPC1 region (ADO79405). The gene DLGI (discs,
CC   large homolog 1 (Drosophila)) is also known as synapse-associated protein
CC   97, hdlg or SAP97. DLG1 has been mapped to chromosomal position 3q29. The
CC   gene KIAA0783 is also known as PHF14 and PHD finger protein 14. KIAA0783
CC   has been mapped to chromosomal position 7p21.3. The KIAA0783 protein is a
CC   novel gene with unknown function, however, being a zinc finger protein,
CC   it likely to be a transcription factor. The gene DPF3 (D4, zinc and
CC   double PHD fingers, family 3) is also known as CERD4, cer-d4, FLJ14079
CC   and 2810403B03Rik. DPF3 is a Rho family guanine-nucleotide exchange
CC   factor. DPF3 has been mapped to chromosomal position 14q24.3-q31.1. The
CC   gene CENPCI (centromere protein C1) is also known as Centromere
CC   autoantigen C1. CENPC1 has been mapped to chromosomal position 4q12-
CC   q13.3. CENPC1 is a centromere autoantigen and a component of the inner
CC   kinetochore plate. The CENPC1 protein is required for maintaining proper
CC   kinetochore size and a timely transition to anaphase. The method is
CC   useful for identifying a subject at risk of breast cancer, for early
CC   diagnosis, prevention and treatment of breast cancer, to analyze and
CC   predict a response to a breast cancer treatment, and in clinical drug
CC   trials.
XX
SQ   Sequence 76600 BP; 25920 A; 14119 C; 13638 G; 22902 T; 0 U; 21 Other;

    Query Match   100.0%; Score 600; DB 12; Length 76600;
    Best Local Similarity   99.8%; Pred. No. 2.4e-84;
    Matches 599; Conservative   1; Mismatches   0; Indels   0; Gaps   0;

Qy   1 TACAATAGACCCTGCTTCTTTCAACGCTTCAACTGCTTTGCTATGTGTTACATCACGAAC 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   39901 TACAATAGACCCTGCTTCTTTCAACGCTTCAACTGCTTTGCTATGTGTTACATCACGAAC 39960

Qy   61 ATCTACTTCATTTACTTGTAATATACAGTCATTGACCCTGAGAAAACAATTAAATATTAA 120
              ||||||||||||||||:|||||||||||||||||||||||||||||||||||||||||||||
Db   39961 ATCTACTTCATTTACTYGTAATATACAGTCATTGACCCTGAGAAAACAATTAAATATTAA 40020

Qy   121 AAATAAAAATTAAATATAATTTAAATATAAATTTTCCACTTTACCAATTTTTTGTTACTT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   40021 AAATAAAAATTAAATATAATTTAAATATAAATTTTCCACTTTACCAATTTTTTGTTACTT 40080

Qy   181 CTTTTTTAAGGTAAAGAGAATTATAAATAATTCTGGAGTAATTCCAGAAAACATAAATGA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   40081 CTTTTTTAAGGTAAAGAGAATTATAAATAATTCTGGAGTAATTCCAGAAAACATAAATGA 40140

Qy   241 AGAAAGTATATCAAAAACTAATATAAACAAATACAAACATTTCCCAAGGGCCAGCAAAAG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   40141 AGAAAGTATATCAAAAACTAATATAAACAAATACAAACATTTCCCAAGGGCCAGCAAAAG 40200

Qy   301 GAACAAAAGAAATAGTGAAAATAATAGATTATATAAAAATGTTAAATAATAATTACAGCT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   40201 GAACAAAAGAAATAGTGAAAATAATAGATTATATAAAAATGTTAAATAATAATTACAGCT 40260

Qy   361 ACCATTTGCATACTTCCTGCATGCTTATTCTGTGTAGGCATTAAGACACACTTTATAAAA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   40261 ACCATTTGCATACTTCCTGCATGCTTATTCTGTGTAGGCATTAAGACACACTTTATAAAA 40320

```
Qy        421 ATAGCAAACATTTATTTAGCACTAACCACATGCCAGGCACTTTCTTGGTATTTTAACCCT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      40321 ATAGCAAACATTTATTTAGCACTAACCACATGCCAGGCACTTTCTTGGTATTTTAACCCT 40380

Qy        481 CATGACACCTGTAAGCTTAATATATATTTTAATCCCTATTTCACAGATGGAGAAACTGAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      40381 CATGACACCTGTAAGCTTAATATATATTTTAATCCCTATTTCACAGATGGAGAAACTGAG 40440

Qy        541 GCACAAAGAATGTAAATAACTTTCCTAAGGCCACCCAGATAATAAGTGACAGAGCTGTGA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      40441 GCACAAAGAATGTAAATAACTTTCCTAAGGCCACCCAGATAATAAGTGACAGAGCTGTGA 40500
```

RESULT 2
AEE84582/c
```
ID   AEE84582 standard; DNA; 2778 BP.
XX
AC   AEE84582;
XX
DT   23-FEB-2006  (first entry)
XX
DE   Discs large homolog 1 (DLG1) gene.
XX
KW   ds; gene; Discs large homolog 1; DLG1; pharmaceutical; diagnostic;
KW   screening; apoptosis; cancer; neurodegenerative disease; breast tumor;
KW   cytostatic; endocrine disease; gynecology and obstetrics; neoplasm;
KW   lung tumor; respiratory disease; Alzheimers disease; neuroprotective;
KW   nootropic; degeneration; neurological disease.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .2778
FT                   /*tag= a
FT                   /product= "Discs large homolog 1 (DLG1)"
XX
PN   WO2005118631-A1.
XX
PD   15-DEC-2005.
XX
PF   02-JUN-2005; 2005WO-JP010538.
XX
PR   03-JUN-2004; 2004JP-00165407.
XX
PA   (TAKE ) TAKEDA PHARM CO LTD.
XX
PI   Sunahara E,  Ishii T;
XX
DR   WPI; 2006-047512/05.
DR   P-PSDB; AEE84581.
XX
PT   Complex useful for screening compounds utilized for promoting or
PT   inhibiting apoptosis for treating cancer or neurodegenerative diseases,
PT   comprises semaphorin 4B protein and discs large homolog 1 and/or 3
PT   protein.
XX
```

XX

CC   The invention relates to a complex (I) comprising semaphorin 4B (SEMA4B)
CC   protein (P1) having an amino acid sequence which is similar or
CC   substantially similar to that of (SEQ ID N0:26; see AEE84580), and discs
CC   large homolog 1 (DLG1) and/or DLG3 (P2) having an amino acid sequence
CC   which is similar or substantially similar to that of (SEQ ID N0:27 or SEQ
CC   ID NO:29; see AEE84581and AEE84583). (P1) of (I) further comprises one of
CC   four fully amino acid sequences (SEQ ID No. 1, 4, 7 or 10; see AEE84555,
CC   AEE84558, AEE84561 and AEE84564). Also included are: an antibody (II)
CC   with respect to (I), or that inhibits or promotes dissociation or
CC   formation of (I); a pharmaceutical (III) comprising (II); a diagnostic
CC   (IV) comprising (II); screening (M1) compound or its salt that inhibits
CC   or promotes binding of (P1) with (P2) or dissociation of (I), involves
CC   utilizing (P1) or (P2); a kit for carrying out (M1), comprising (P1) or
CC   (P2); a compound (V) or its salt that inhibits binding of (P1) with (P2)
CC   or promotes dissociation of (I); a promoter (VI) of apoptosis of cancer
CC   cell, the cancer proliferation inhibitor (VII) or preventive/therapeutic
CC   agent of cancer (VIII), comprising (V) or its salt; a compound (IX) or
CC   its salt that promotes binding of (P1) with (P2) or inhibits dissociation
CC   of (I); an inhibitor (X) of apoptosis of neuron, or
CC   preventive/therapeutic agent (XI) of neurodegenerative diseases,
CC   comprising (IX) or its salt; promoting (M2) apoptosis or inhibiting (M3)
CC   cancer cell proliferation, involves inhibiting binding of (P1) and (P2)
CC   or promoting dissociation of (I); preventing or treating (M4) cancer,
CC   involves inhibiting binding of (P1) and (P2) or promoting dissociation of
CC   (I); inhibiting (M5) apoptosis of neuron or preventing or treating (M6)
CC   neurodegenerative diseases, involves promoting binding of (P1) and (P2)
CC   or inhibiting dissociation of (I); screening compound or its salt for
CC   preventing or treating cancer or neurodegenerative diseases, involves
CC   utilizing (P2), its partial peptide or its salt; the use of a substance
CC   (XII) or its salt that inhibits binding of (P1) and (P2) or promotes
CC   dissociation of (I) for producing promoter of apoptosis of cancer cell,
CC   cancer cell proliferation inhibitor or preventive/therapeutic agent of
CC   cancer; and the use of a substance (XIII) or its salt that promotes
CC   binding of (P1) and (P2) or inhibits dissociation of (I) for producing
CC   inhibitor of apoptosis of neuron or preventive/therapeutic agent of
CC   neurodegenerative diseases. (III) is useful as promoter of apoptosis of
CC   cancer cell, cancer cell proliferation inhibitor or
CC   preventive/therapeutic agent of cancer, or as apoptosis inhibitor of
CC   neuron or preventive/therapeutic agent of neurodegenerative diseases.
CC   (IV) is useful for diagnosing cancer or neurodegenerative diseases. (VI),
CC   (VII) or (VIII) is useful for promoting apoptosis of cancer cell,
CC   inhibiting cancer proliferation, or preventing or treating cancer. (X) or
CC   (XI) is useful for inhibiting apoptosis of neuron, or preventing or
CC   treating neurodegenerative diseases. (M2) is useful for promoting
CC   apoptosis. (M3) is useful for inhibiting cancer cell proliferation. (M4)
CC   is useful for preventing or treating cancer. (M5) is useful for
CC   inhibiting apoptosis of neurons. (M6) is useful for preventing or
CC   treating neurodegenerative diseases. (XII) is useful for producing
CC   promoter of apoptosis of cancer cell, cancer cell proliferation inhibitor
CC   or preventive/therapeutic agent of cancer. (XIII) is useful for producing
CC   inhibitor of apoptosis of neuron or preventive/therapeutic agent of
CC   neurodegenerative diseases. (I) is useful for screening compounds
CC   utilized for promoting or inhibiting apoptosis for treating cancer (e.g.
CC   breast or lung cancer) or neurodegenerative diseases (e.g. Alzheimer's
CC   disease). (I) enables screening of promoter or inhibitor of apoptosis.

CC   The present sequence represents Discs large homolog 1 (DLG1) gene.
XX
SQ   Sequence 2778 BP; 930 A; 545 C; 605 G; 698 T; 0 U; 0 Other;

     Query Match   16.3%;   Score 98;   DB 15;   Length 2778;
     Best Local Similarity 100.0%; Pred. No. 1.5e-06;
     Matches 98; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   1 TACAATAGACCCTGCTTCTTTCAACGCTTCAACTGCTTTGCTATGTGTTACATCACGAAC 60
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   909 TACAATAGACCCTGCTTCTTTCAACGCTTCAACTGCTTTGCTATGTGTTACATCACGAAC 850

Qy   61 ATCTACTTCATTTACTTGTAATATACAGTCATTGACCC 98
             ||||||||||||||||||||||||||||||||||||||
Db   849 ATCTACTTCATTTACTTGTAATATACAGTCATTGACCC 812

RESULT 3
AED21168/c
ID   AED21168 standard; DNA; 2980 BP.
XX
AC   AED21168;
XX
DT   15-DEC-2005   (first entry)
XX
DE   Human Dlg (disks large) DNA sequence.
XX
KW   secreted frizzled-related protein; tumor; cytostatic; neoplasm;
KW   drug screening; skin tumor; lymphoma; hematological disease;
KW   immune disorder; disks large; Dlg; ds; gene.
XX
OS   Homo sapiens.
XX
FH   Key   Location/Qualifiers
FT   CDS   189. .2903
FT   /*tag= a
FT   /product= "disks large protein"
XX
PN   WO2005094887-A1.
XX
PD   13-OCT-2005.
XX
PF   30-MAR-2005; 2005WO-JP006163.
XX
PR   31-MAR-2004; 2004JP-00106315.
XX
PA   (DAUC) DAIICHI PHARM CO LTD.
XX
PI   Akiyama T, Ishidao T, Aiba T;
XX
DR   WPI; 2005-725415/74.
DR   P-PSDB; AED21169.
XX
PT   Enhancer of expression and/or function of secreted frizzled-related
PT   protein for inhibiting tumorigenesis, comprises a compound capable of
PT   increasing expression and/or function of discs large.
XX

PS   Disclosure; SEQ ID NO 1; 54pp; Japanese.
XX
CC   The new invention relates to the finding that tumor formation and the
CC   transcription of secreted frizzled-related protein (sFRP) gene is lowered
CC   in a Dlg-knockout mouse. Specifically claimed is an enhancer (I) of the
CC   expression and/or function of secreted frizzled-related protein (sFRP),
CC   comprising a compound capable of increasing the expression and/or
CC   function of Dlg (discs large). Also claimed are a tumorigenesis inhibitor
CC   (II) containing (I); an agent (III) for preventing or treating tumor,
CC   containing (I); increasing (M1) the expression and/or function of sFRP,
CC   comprising inducing the expression and/or function of Dlg; a non-human
CC   mammal (IV) having one of the deleted Dlg allele; a cell (V) derived from
CC   a non-human mammal having a deleted Dlg allele; and detecting tumor
CC   tissue or cells, comprising measuring the expression and/or function of
CC   Dlg in a sample of the cell or tissue, comparing the expression and/or
CC   function with a healthy tissue or a cell and detecting whether the
CC   expression and/or function is reduced. In (I), the compound that
CC   increases the expression and/or function of Dlg is chosen from a Dlg gene
CC   or recombinant vector containing the Dlg gene. The sFRP is sFRP2. (M1)
CC   involves using a Dlg gene or recombinant vector containing the Dlg gene.
CC   (I) or (M1) are useful for inhibiting tumorigenesis and for preventing or
CC   treating tumor. (IV) or (V) is useful for identifying a compound capable
CC   of increasing the expression and/or function of Dlg or sFRP, or
CC   inhibiting tumorigenesis. (II) and (III) are useful for treating a tumor
CC   such as skin tumor or lymphoma. (I) effectively induces the expression
CC   and/or function of sFRP. The present sequence is human disks large DNA
CC   sequence.
XX
SQ   Sequence 2980 BP; 949 A; 595 C; 672 G; 764 T; 0 U; 0 Other;

      Query Match   16.3%;   Score 98;   DB 14;   Length 2980;
      Best Local Similarity   100.0%;   Pred. No. 1.5e-06;
      Matches   98;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

Qy   1 TACAATAGACCCTGCTTCTTTCAACGCTTCAACTGCTTTGCTATGTGTTACATCACGAAC 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1097 TACAATAGACCCTGCTTCTTTCAACGCTTCAACTGCTTTGCTATGTGTTACATCACGAAC 1038

Qy   61 ATCTACTTCATTTACTTGTAATATACAGTCATTGACCC 98
              ||||||||||||||||||||||||||||||||||||||
Db   1037 ATCTACTTCATTTACTTGTAATATACAGTCATTGACCC 1000

RESULT 4
AEF75023/c
ID   AEF75023 standard; DNA; 2980 BP.
XX
AC   AEF75023;
XX
DT   06-APR-2006   (first entry)
XX
DE   Human polynucleotide #537.
XX
KW   Diagnosis; gene regulation; gene expression;
KW   post traumatic stress disorder; psychiatric disorder; tranquilizer; gene;
KW   ds.
XX

```
OS    Homo sapiens.
XX
PN    WO2006013561-A2.
XX
PD    09-FEB-2006.
XX
PF    02-AUG-2005; 2005WO-IL000824.
XX
PR    02-AUG-2004; 2004US-0592408P.
XX
PA    (YISS ) YISSUM RES DEV CO HEBREW UNIV JERUSALEM.
PA    (HADA-) HADASIT MEDICAL RES SERVICES & DEV LTD.
XX
PI    Segman R,  Shalev A,  Goltser T,  Friedman N,  Shefi N,  Kaminski N;
XX
DR    WPI; 2006-145797/15.
XX
PT    New kit comprising 10 and no more than 574 polynucleotides capable of
PT    specifically binding at least one specific polynucleotide sequence,
PT    useful for determining predisposition of a subject to develop PTSD, or
PT    for diagnosing PTSD.
XX
PS    Claim 1; SEQ ID NO 537; 157pp; English.
XX
CC    The invention relates to a kit for determining predisposition of a
CC    subject to developing post-traumatic stress disorder (PTSD) comprising at
CC    least 10 and no more than 574 polynucleotides, where each of the
CC    polynucleotides is capable of specifically binding at least one specific
CC    polynucleotide sequence. The invention also relates to a kit for
CC    diagnosing PTSD in a subject, agents for the manufacture of the kits
CC    cited comprising the polynucleotides cited, and a microarray comprising
CC    at least 10 and no more than 904 oligonucleotides where each of the
CC    oligonucleotides is capable of specifically binding at least one specific
CC    polynucleotide sequence. The kit comprises each of the polynucleotides
CC    selected from an oligonucleotide molecule, a cDNA molecule, a genomic
CC    molecule and an RNA molecule. Each of the polynucleotides is at least 10
CC    and no more than 50 nucleic acids in length. Each of the polynucleotides
CC    is bound to a solid support. The kit also comprises at least one reagent
CC    suitable for detecting hybridization of the polynucleotides and at least
CC    one RNA transcript. The kit further comprises packaging materials
CC    packaging the at least one reagent and instructions for using the kit in
CC    determining predisposition of the subject to developing PTSD, or for
CC    diagnosing the disease. The microarray comprises oligonucleotides of at
CC    least 10 and no more than 40 nucleic acids in length. The agent is
CC    capable of regulating an expression level of at least one gene as a
CC    pharmaceutical or for the manufacture of a medicament identified for
CC    preventing PTSD. The kit is useful for determining predisposition of a
CC    subject to developing PTSD or for diagnosing PTSD. This sequence
CC    represents a human polynucleotide of the invention. Note: The sequence
CC    data for this patent did not form part of the printed specification, but
CC    was obtained in electronic format directly from WIPO at
CC    ftp.wipo.int/pub/published_pct_sequences.
XX
SQ    Sequence 2980 BP; 949 A; 595 C; 672 G; 764 T; 0 U; 0 Other;

      Query Match              16.3%;  Score 98;  DB 15;  Length 2980;
      Best Local Similarity    100.0%;  Pred. No. 1.5e-06;
```

```
       Matches   98;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         1 TACAATAGACCCTGCTTCTTTCAACGCTTCAACTGCTTTGCTATGTGTTACATCACGAAC 60
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1097 TACAATAGACCCTGCTTCTTTCAACGCTTCAACTGCTTTGCTATGTGTTACATCACGAAC 1038

Qy        61 ATCTACTTCATTTACTTGTAATATACAGTCATTGACCC 98
             ||||||||||||||||||||||||||||||||||||||
Db      1037 ATCTACTTCATTTACTTGTAATATACAGTCATTGACCC 1000
```

RESULT 5 ε
ACC85071/c
```
ID    ACC85071 standard; DNA; 3046 BP.
XX
AC    ACC85071;
XX
DT    13-OCT-2003   (first entry)
XX
DE    Human MBCAT polypeptide encoding DNA.
XX
KW    MBCAT; beta-catenin; cytostatic; gene therapy; cancer; human; gene; ds.
XX
OS    Homo sapiens.
XX
FH    Key              Location/Qualifiers
FT    CDS              189. .2969
FT                     /*tag=  a
FT                     /product= "MBCAT"
XX
PN    WO2003052068-A2.
XX
PD    26-JUN-2003.
XX
PF    12-DEC-2002; 2002WO-US039796.
XX
PR    13-DEC-2001; 2001US-0340213P.
PR    13-DEC-2001; 2001US-0340314P.
PR    13-DEC-2001; 2001US-0340322P.
PR    15-FEB-2002; 2002US-0357502P.
XX
PA    (EXEL-) EXELIXIS INC.
XX
PI    Costa MA,  Gendreau SB,  Dora EG,  Nicoll M;
XX
DR    WPI; 2003-533010/50.
DR    P-PSDB; ABR82220.
XX
PT    Identifying a candidate beta-catenin pathway modulating agent for
PT    diagnosing or treating cancer by detecting a test agent-biased activity
PT    of the assay system comprising a purified MBCAT polypeptide or nucleic
PT    acid.
XX
PS    Example; Page 49-51; 81pp; English.
XX
CC    The invention relates to genes that modify beta-catenin pathway and to
CC    the identification of human MBCAT (modifiers of beta-catenin) genes. The
```

```
CC    MBCAT polypeptides are therapeutic targets for disorders associated with
CC    beta-catenin function and are useful for manufacturing a medicament for
CC    diagnosing or treating breast, colon, lung or ovary cancer. The present
CC    sequence represents a human MBCAT polypeptide encoding DNA
XX
SQ    Sequence 3046 BP; 975 A; 612 C; 681 G; 778 T; 0 U; 0 Other;

      Query Match             16.3%;  Score 98;  DB 9;  Length 3046;
      Best Local Similarity   100.0%;  Pred. No. 1.5e-06;
      Matches   98;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         1 TACAATAGACCCTGCTTCTTTCAACGCTTCAACTGCTTTGCTATGTGTTACATCACGAAC 60
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1097 TACAATAGACCCTGCTTCTTTCAACGCTTCAACTGCTTTGCTATGTGTTACATCACGAAC 1038

Qy        61 ATCTACTTCATTTACTTGTAATATACAGTCATTGACCC 98
             ||||||||||||||||||||||||||||||||||||||
Db      1037 ATCTACTTCATTTACTTGTAATATACAGTCATTGACCC 1000
```

RESULT 6
ADK11476/c

```
ID    ADK11476 standard; DNA; 3046 BP.
XX
AC    ADK11476;
XX
DT    06-MAY-2004  (first entry)
XX
DE    Human discs large (Drosophila) homolog 1 gene.
XX
KW    ds; gene; cytostatic; cardiovascular; immunosuppressive; nephrotropic;
KW    antirheumatic; antiarthritic; dermatological; antipsoriatic;
KW    antiinflammatory; fungicide; gene therapy; Drosophila; diagnosis;
KW    cardiovascular disorder; autoimmune disease; glomerulonephritis;
KW    rheumatoid arthritis; dermatological disorder; psoriasis;
KW    inflammatory disorder; malaria; emphysema; alopecia.
XX
OS    Homo sapiens.
XX
PN    WO2003040301-A2.
XX
PD    15-MAY-2003.
XX
PF    23-OCT-2002; 2002WO-GB004780.
XX
PR    05-NOV-2001; 2001GB-00026506.
PR    27-NOV-2001; 2001GB-00028384.
PR    11-FEB-2002; 2002GB-00003185.
XX
PA    (CYCL-) CYCLACEL LTD.
XX
PI    Deak P, Frenz L, Glover D, Midgley C;
XX
DR    WPI; 2003-441540/41.
DR    P-PSDB; ADK11477.
XX
PT    New Drosophila polypeptides and polynucleotides, useful for diagnosing,
```

PT   preventing and/or treating disorders, such as cancer, glomerulonephritis,
PT   rheumatoid arthritis, psoriasis, malaria, emphysema and alopecia.
XX
PS   Example 28; Page 225; 265pp; English.
XX
CC   The invention relates to novel Drosophila species DNA sequences and their
CC   encoded proteins with their corresponding human homologues. The proteins
CC   or their encoding polynucleotides are useful in a method of prevention,
CC   treatment or diagnosis of a disease in an individual, and used to
CC   identify a substance capable of binding to the polypeptide or modulating
CC   the function of the polypeptide comprising incubating the polypeptide
CC   with a candidate substance and determining whether the substance binds to
CC   the polypeptide. The compositions are administered to an individual in
CC   need of such treatment. The method of diagnosis, in which the presence or
CC   absence of a polynucleotide is detected in a biological sample, comprises
CC   brining the biological sample containing the nucleic acid such as DNA or
CC   RNA into contact with a probe comprising a fragment of at least 15
CC   nucleotides of the polynucleotide, and detecting any duplex formed
CC   between the probe and nucleic acid in the sample. The method also
CC   comprises providing an antibody capable of binding to the polypeptide,
CC   incubating a biological sample with the antibody to allow the formation
CC   of an antibody-antigen complex, and determining whether antibody-antigen
CC   complex comprising the antibody is formed. The disease comprises a
CC   proliferative disease such as cancer. The antibody or identified
CC   substance is also useful in inhibiting the function of a polypeptide
CC   and/or regulating a cell division cycle function. The diseases also
CC   include cardiovascular disorders, autoimmune diseases such as
CC   glomerulonephritis and rheumatoid arthritis, and dermatological disorders
CC   such as psoriasis, inflammatory, fungal, and parasitic disorders such as
CC   malaria, emphysema and alopecia. This sequence represents a human homolog
CC   gene for one of the Drosophila genes of the invention.
XX
SQ   Sequence 3046 BP; 975 A; 612 C; 681 G; 778 T; 0 U; 0 Other;

     Query Match   16.3%; Score 98; DB 10; Length 3046;
     Best Local Similarity   100.0%; Pred. No. 1.5e-06;
     Matches   98; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 TACAATAGACCCTGCTTCTTTCAACGCTTCAACTGCTTTGCTATGTGTTACATCACGAAC 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1097 TACAATAGACCCTGCTTCTTTCAACGCTTCAACTGCTTTGCTATGTGTTACATCACGAAC 1038

Qy   61 ATCTACTTCATTTACTTGTAATATACAGTCATTGACCC 98
              ||||||||||||||||||||||||||||||||||||||
Db   1037 ATCTACTTCATTTACTTGTAATATACAGTCATTGACCC 1000

RESULT 7
ADR83538/c
ID   ADR83538 standard; DNA; 3046 BP.
XX
AC   ADR83538;
XX
DT   02-DEC-2004   (first entry)
XX
DE   Human discs large homolog 1 DNA, target gene of miRNA.
XX

CC   This invention relates to a novel method for modulating the expression of
CC   a target gene in a cell. Specifically, it refers to the introduction into
CC   a cell of a polynucleotide that forms a duplex region with an mRNA
CC   transcribed from the target gene, where the duplex region comprises a
CC   mammalian miRNA target region i.e. a non-coding microRNA (miRNA) that
CC   regulates mRNA at a post-transcriptional level. The present invention
CC   describes a method for controlling ontogenesis of a mammal, function of a
CC   mammalian cell, differentiation of a mammalian cell or viability of a
CC   mammalian cell in the post-transcriptional phase, which comprises
CC   introducing a plasmid vector comprising a promoter and nucleic acid
CC   molecule expressing an miRNA or siRNA silencing precursor to the miRNA.
CC   Accordingly, it provides a cell therapy method for treating cancer,
CC   immune disease, nerve disorder (e.g. amyotrophic lateral sclerosis,
CC   Parkinson's disease, or Alzheimer's disease) or an inflammatory disease
CC   by introducing into the cell the miRNA, siRNA silencing precursor to the
CC   miRNA or the plasmid vector. As such, they can be developed into
CC   pharmaceutical compositions that exhibit cytostatic, immunosuppressive,
CC   nootropic, neuroprotective and antiinflammatory activities and hence can
CC   be used for immunotherapy. This polynucleotide sequence is a human target
CC   gene whose expression is modulated by miRNAs of the invention.
XX
SQ   Sequence 3046 BP; 975 A; 612 C; 681 G; 778 T; 0 U; 0 Other;

     Query Match   16.3%;   Score 98;   DB 13;   Length 3046;
     Best Local Similarity   100.0%;   Pred. No. 1.5e-06;
     Matches   98; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 TACAATAGACCCTGCTTCTTTCAACGCTTCAACTGCTTTGCTATGTGTTACATCACGAAC 60

```
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1097 TACAATAGACCCTGCTTCTTTCAACGCTTCAACTGCTTTGCTATGTGTTACATCACGAAC 1038

Qy     61 ATCTACTTCATTTACTTGTAATATACAGTCATTGACCC 98
          ||||||||||||||||||||||||||||||||||||||
Db   1037 ATCTACTTCATTTACTTGTAATATACAGTCATTGACCC 1000
```

RESULT 8
ADU05806/c

ID	ADU05806 standard; DNA; 3046 BP.
XX	
AC	ADU05806;
XX	
DT	27-JAN-2005 (first entry)
XX	
DE	Novel bronchial cancer-associated human gene SeqID28.
XX	
KW	bronchial cancer; cytostatic; tumour-associated protein;
KW	cancer detection; metastasis; tumour; gene; ds; human.
XX	
OS	Homo sapiens.
XX	
PN	DE10316701-A1.
XX	
PD	04-NOV-2004.
XX	
PF	09-APR-2003; 2003DE-01016701.
XX	
PR	09-APR-2003; 2003DE-01016701.
XX	
PA	(HINZ/) HINZMANN B.
PA	(HERM/) HERMANN K.
PA	(CAST/) HEIDEN CASTANOS-VELEZ E.
XX	
PI	Mennerich D, Bruemmendorf T, Heiden E, Hermann K, Kinnemann H;
PI	Li X, Roepcke S, Staub E, Hinzmann B, Rosenthal A, Pilarsky C;
XX	
DR	WPI; 2004-786403/78.
DR	P-PSDB; ADU06293.
XX	
PT	New nucleic acid, and derived proteins, useful for diagnosis of bronchial
PT	cancer and in screening for therapeutic and diagnostic agents.
XX	
PS	Claim 1; SEQ ID NO 28; 1381pp; German.
XX	
CC	This invention relates to a novel isolated nucleic acid associated with
CC	bronchial cancer comprising 489 defined sequences given in the
CC	specification. The invention may be useful for the production of
CC	compounds with a cytostatic activity through the inhibition of expression
CC	or activity of tumour-associated proteins. The novel DNA sequences and
CC	the proteins/peptides encoded by them are used for detecting bronchial
CC	cancer or determining the risk of developing it and to screen for
CC	specific binding partners of the DNA or protein sequences, where the
CC	binding partners are potentially useful as agents for treating or
CC	diagnosing bronchial cancer. The DNA or protein sequences can also be
CC	used for prognosis, detection of metastases and for secondary treatment

OM nucleic - nucleic search, using sw model

Run on: July 15, 2006, 18:16:32 ; Search time 1201 Seconds
                                         (without alignments)
                                         6138.702 Million cell updates/sec

Title: US-10-722-939-1-T39977_39901-40500
Perfect score: 600
Sequence: 1 tacaatagaccctgcttctt.........ataagtgacagagctgtga 600

Scoring table: IDENTITY_NUC
                 Gapop 10.0 , Gapext 1.0

Searched: 18892170 seqs, 6143817638 residues

Total number of hits satisfying chosen parameters: 37784340

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 100 summaries

Database :   Published_Applications_NA_Main:*
                 1: /EMC_Celerra_SIDS3/ptodata/2/pubpna/US07_PUBCOMB.seq:*
                 2: /EMC_Celerra_SIDS3/ptodata/2/pubpna/US08_PUBCOMB.seq:*
                 3: /EMC_Celerra_SIDS3/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
                 4: /EMC_Celerra_SIDS3/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
                 5: /EMC_Celerra_SIDS3/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
                 6: /EMC_Celerra_SIDS3/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
                 7: /EMC_Celerra_SIDS3/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
                 8: /EMC_Celerra_SIDS3/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
                 9: /EMC_Celerra_SIDS3/ptodata/2/pubpna/US10D_PUBCOMB.seq:*
                 10: /EMC_Celerra_SIDS3/ptodata/2/pubpna/US10E_PUBCOMB.seq:*
                 11: /EMC_Celerra_SIDS3/ptodata/2/pubpna/US10F_PUBCOMB.seq:*
                 12: /EMC_Celerra_SIDS3/ptodata/2/pubpna/US10G_PUBCOMB.seq:*
                 13: /EMC_Celerra_SIDS3/ptodata/2/pubpna/US11A_PUBCOMB.seq:*
                 14: /EMC_Celerra_SIDS3/ptodata/2/pubpna/US11B_PUBCOMB.seq:*
                 15: /EMC_Celerra_SIDS3/ptodata/2/pubpna/US11C_PUBCOMB.seq:*
                 16: /EMC_Celerra_SIDS3/ptodata/2/pubpna/US11D_PUBCOMB.seq:*

         Pred. No. is the number of results predicted by chance to have a
         score greater than or equal to the score of the result being printed,
         and is derived by analysis of the total score distribution.

                              SUMMARIES

                      %
Result   Query
   No.   Score Match Length DB   ID   Description
----
     1   599.6 100.0 76600 10 US-10-722-939-1   Sequence 1, Appli

	2	543.6	90.6	591	4	US-09-925-065A-649342	Sequence 649342,
	3	543.6	90.6	591	5	US-09-925-065A-649342	Sequence 649342,
c	4	98	16.3	3046	10	US-10-840-060-266	Sequence 266, App
c	5	97.6	16.3	3046	10	US-10-722-939-5	Sequence 5, Appli
c	6	96	16.0	1152	6	US-10-106-698-1545	Sequence 1545, Ap
c	7	82.6	13.8	419	3	US-09-960-352-11234	Sequence 11234, A
c	8	76.8	12.8	8056	9	US-10-473-126-386	Sequence 386, App
c	9	76.6	12.8	542	9	US-10-425-115-7873	Sequence 7873, Ap
c	10	76.2	12.7	3256	10	US-10-486-706-306	Sequence 306, App
c	11	76.2	12.7	3256	10	US-10-980-571-14	Sequence 14, Appl
c	12	75.2	12.5	8056	9	US-10-473-126-240	Sequence 240, App
c	13	74.8	12.5	516	3	US-09-960-352-5785	Sequence 5785, Ap
c	14	73.2	12.2	6145	7	US-10-311-455-945	Sequence 945, App
c	15	72.8	12.1	6292	8	US-10-221-714A-461	Sequence 461, App
c	16	72.6	12.1	14615	8	US-10-221-714A-429	Sequence 429, App
c	17	71.2	11.9	5845	7	US-10-311-455-1635	Sequence 1635, Ap
c	18	71.2	11.9	6161	7	US-10-311-455-384	Sequence 384, App
c	19	71	11.8	363	3	US-09-814-353-4589	Sequence 4589, Ap
c	20	71	11.8	363	3	US-09-814-353-10888	Sequence 10888, A
c	21	70.4	11.7	547	8	US-10-021-323-16890	Sequence 16890, A
c	22	70.4	11.7	513509	3	US-09-754-853A-4	Sequence 4, Appli
c	23	70.2	11.7	16766	7	US-10-311-455-2130	Sequence 2130, Ap
c	24	70	11.7	3683	9	US-10-473-126-339	Sequence 339, App
c	25	70	11.7	12237	7	US-10-311-455-2331	Sequence 2331, Ap
c	26	69.8	11.6	83391	8	US-10-433-793-123	Sequence 123, App
c	27	69.6	11.6	849	8	US-10-424-599-142110	Sequence 142110,
	28	69.6	11.6	3413	8	US-10-437-963-64459	Sequence 64459, A
c	29	69.6	11.6	7442	8	US-10-221-714A-409	Sequence 409, App
c	30	69.2	11.5	408	3	US-09-960-352-1221	Sequence 1221, Ap
c	31	69.2	11.5	5413	7	US-10-311-455-537	Sequence 537, App
c	32	69	11.5	5750	7	US-10-311-455-1981	Sequence 1981, Ap
c	33	69	11.5	5750	8	US-10-221-714A-433	Sequence 433, App
c	34	69	11.5	6154	8	US-10-221-613-69	Sequence 69, Appl
c	35	69	11.5	3673778	7	US-10-312-841-2	Sequence 2, Appli
c	36	68.8	11.5	5572	7	US-10-311-455-1398	Sequence 1398, Ap
c	37	68.8	11.5	5572	7	US-10-240-452-60	Sequence 60, Appl
c	38	68.8	11.5	7657	6	US-10-239-676-185	Sequence 185, App
c	39	68.8	11.5	7657	7	US-10-311-455-1995	Sequence 1995, Ap
	40	68.6	11.4	1214	8	US-10-424-599-102083	Sequence 102083,
c	41	68.6	11.4	40324	8	US-10-433-793-179	Sequence 179, App
	42	68.6	11.4	48933	6	US-10-087-192-838	Sequence 838, App
	43	68.2	11.4	8056	9	US-10-473-126-240	Sequence 240, App
c	44	68	11.3	3673778	7	US-10-312-841-1	Sequence 1, Appli
c	45	67.8	11.3	393	3	US-09-960-352-4582	Sequence 4582, Ap
c	46	67.8	11.3	12405	6	US-10-239-676-35	Sequence 35, Appl
c	47	67.8	11.3	12405	7	US-10-240-453-43	Sequence 43, Appl
c	48	67.8	11.3	12405	8	US-10-221-613-101	Sequence 101, App
c	49	67.6	11.3	314	3	US-09-960-352-12412	Sequence 12412, A
c	50	67.6	11.3	673	8	US-10-424-599-57885	Sequence 57885, A
c	51	67.6	11.3	6775	8	US-10-433-793-189	Sequence 189, App
	52	67.6	11.3	8056	9	US-10-473-126-386	Sequence 386, App
c	53	67.6	11.3	26997	8	US-10-221-714A-473	Sequence 473, App
c	54	67.4	11.2	10891	7	US-10-311-455-437	Sequence 437, App
c	55	67.4	11.2	11745	7	US-10-240-453-205	Sequence 205, App
c	56	67.2	11.2	5406	7	US-10-311-455-1071	Sequence 1071, Ap
c	57	67.2	11.2	7321	8	US-10-221-613-300	Sequence 300, App
c	58	67.2	11.2	15548	7	US-10-311-455-2128	Sequence 2128, Ap

```
c  59    67     11.2      495   8  US-10-021-323-2679      Sequence 2679, Ap
c  60    67     11.2     6552   8  US-10-221-714A-400      Sequence 400, App
c  61    67     11.2     6767   8  US-10-221-714A-330      Sequence 330, App
c  62    67     11.2    14987   7  US-10-311-455-604       Sequence 604, App
   63    67     11.2    94330   6  US-10-087-192-1222      Sequence 1222, Ap
c  64    66.8   11.1      262   3  US-09-854-867-99        Sequence 99, Appl
c  65    66.8   11.1      262  10  US-10-786-970A-99        Sequence 99, Appl
c  66    66.8   11.1      446   3  US-09-960-352-3400      Sequence 3400, Ap
c  67    66.8   11.1      560   8  US-10-021-323-2253      Sequence 2253, Ap
c  68    66.8   11.1     6171   7  US-10-311-455-761       Sequence 761, App
c  69    66.8   11.1   113515   7  US-10-311-455-2148      Sequence 2148, Ap
c  70    66.6   11.1      591   8  US-10-437-963-21828     Sequence 21828, A
   71    66.6   11.1      663   6  US-10-027-632-232886    Sequence 232886,
   72    66.6   11.1      663   7  US-10-027-632-232886    Sequence 232886,
c  73    66.6   11.1    14006   7  US-10-311-455-1931      Sequence 1931, Ap
   74    66.4   11.1     1130   9  US-10-425-115-132399    Sequence 132399,
c  75    66     11.0      375   9  US-10-357-930-44930     Sequence 44930, A
c  76    65.8   11.0      349   3  US-09-814-353-18039     Sequence 18039, A
c  77    65.8   11.0      578   8  US-10-021-323-7450      Sequence 7450, Ap
c  78    65.8   11.0     2739   8  US-10-433-793-206       Sequence 206, App
c  79    65.8   11.0     3315  10  US-10-486-319A-33        Sequence 33, Appl
c  80    65.8   11.0     3315  10  US-10-486-319A-55        Sequence 55, Appl
c  81    65.8   11.0     3814   9  US-10-473-126-150       Sequence 150, App
c  82    65.8   11.0     3814   9  US-10-473-126-296       Sequence 296, App
c  83    65.8   11.0     5476   6  US-10-239-676-207       Sequence 207, App
c  84    65.8   11.0     5476   7  US-10-311-455-2087      Sequence 2087, Ap
c  85    65.8   11.0     5476  10  US-10-240-708-81         Sequence 81, Appl
c  86    65.8   11.0     5888   7  US-10-240-485-9         Sequence 9, Appli
c  87    65.8   11.0     6109   7  US-10-311-455-299       Sequence 299, App
c  88    65.8   11.0     6109   8  US-10-221-613-33        Sequence 33, Appl
c  89    65.8   11.0     6486   8  US-10-433-793-79        Sequence 79, Appl
c  90    65.8   11.0     6815   6  US-10-239-676-50        Sequence 50, Appl
c  91    65.8   11.0     6815   7  US-10-311-455-644       Sequence 644, App
c  92    65.8   11.0     6815   7  US-10-240-453-50        Sequence 50, Appl
c  93    65.8   11.0     9293   6  US-10-239-676-25        Sequence 25, Appl
c  94    65.8   11.0     9293   7  US-10-240-453-33        Sequence 33, Appl
c  95    65.8   11.0     9293   8  US-10-257-166-55        Sequence 55, Appl
c  96    65.6   10.9      390   3  US-09-960-352-3640      Sequence 3640, Ap
   97    65.6   10.9      647   9  US-10-425-115-40312     Sequence 40312, A
   98    65.6   10.9      739   9  US-10-425-115-55020     Sequence 55020, A
c  99    65.6   10.9     2501   9  US-10-473-126-218       Sequence 218, App
c 100    65.6   10.9     2501   9  US-10-473-126-364       Sequence 364, App
```

ALIGNMENTS

RESULT 1
US-10-722-939-1
; Sequence 1, Application US/10722939
; Publication No. US20050214771A1
; GENERAL INFORMATION:
; APPLICANT: ROTH, RICHARD B.
; APPLICANT: NELSON, MATTHEW ROBERTS
; APPLICANT: KAMMERER, STEFAN M.
; APPLICANT: BRAUN, ANDREAS
; APPLICANT: RENELAND, RIKARD

```
;   TITLE OF INVENTION: METHODS FOR IDENTIFYING RISK OF BREAST CANCER AND
TREATMENTS
;   TITLE OF INVENTION:  THEREOF
;   FILE REFERENCE: SEQ-4071-UT
;   CURRENT APPLICATION NUMBER: US/10/722,939
;   CURRENT FILING DATE:  2003-11-26
;   PRIOR APPLICATION NUMBER: 60/429,136
;   PRIOR FILING DATE: 2002-11-25
;   PRIOR APPLICATION NUMBER: 60/490,234
;   PRIOR FILING DATE: 2003-07-23
;   NUMBER OF SEQ ID NOS: 634
;   SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1
;   LENGTH: 76600
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-722-939-1
```

```
     Query Match              100.0%;  Score 600;  DB 10;  Length 76600;
     Best Local Similarity    99.8%;  Pred. No. 1.2e-90;
     Matches  599;  Conservative    1;  Mismatches   0;  Indels   0;  Gaps    0;

Qy          1 TACAATAGACCCTGCTTCTTTCAACGCTTCAACTGCTTTGCTATGTGTTACATCACGAAC 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39901 TACAATAGACCCTGCTTCTTTCAACGCTTCAACTGCTTTGCTATGTGTTACATCACGAAC 39960

Qy         61 ATCTACTTCATTTACTTGTAATATACAGTCATTGACCCTGAGAAAACAATTAAATATTAA 120
              ||||||||||||||||:|||||||||||||||||||||||||||||||||||||||||||
Db      39961 ATCTACTTCATTTACTYGTAATATACAGTCATTGACCCTGAGAAAACAATTAAATATTAA 40020

Qy        121 AAATAAAAATTAAATATAATTTAAATATAAATTTTCCACTTTACCAATTTTTTGTTACTT 180
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      40021 AAATAAAAATTAAATATAATTTAAATATAAATTTTCCACTTTACCAATTTTTTGTTACTT 40080

Qy        181 CTTTTTTAAGGTAAAGAGAATTATAAATAATTCTGGAGTAATTCCAGAAAACATAAATGA 240
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      40081 CTTTTTTAAGGTAAAGAGAATTATAAATAATTCTGGAGTAATTCCAGAAAACATAAATGA 40140

Qy        241 AGAAAGTATATCAAAAACTAATATAAACAAATACAAACATTTCCCAAGGGCCAGCAAAAG 300
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      40141 AGAAAGTATATCAAAAACTAATATAAACAAATACAAACATTTCCCAAGGGCCAGCAAAAG 40200

Qy        301 GAACAAAAGAAATAGTGAAAATAATAGATTATATAAAAATGTTAAATAATAATTACAGCT 360
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      40201 GAACAAAAGAAATAGTGAAAATAATAGATTATATAAAAATGTTAAATAATAATTACAGCT 40260

Qy        361 ACCATTTGCATACTTCCTGCATGCTTATTCTGTGTAGGCATTAAGACACACTTTATAAAA 420
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      40261 ACCATTTGCATACTTCCTGCATGCTTATTCTGTGTAGGCATTAAGACACACTTTATAAAA 40320

Qy        421 ATAGCAAACATTTATTTAGCACTAACCACATGCCAGGCACTTTCTTGGTATTTTAACCCT 480
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      40321 ATAGCAAACATTTATTTAGCACTAACCACATGCCAGGCACTTTCTTGGTATTTTAACCCT 40380

Qy        481 CATGACACCTGTAAGCTTAATATATATTTTAATCCCTATTTCACAGATGGAGAAACTGAG 540
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      40381 CATGACACCTGTAAGCTTAATATATATTTTAATCCCTATTTCACAGATGGAGAAACTGAG 40440
```

```
Qy        541 GCACAAAGAATGTAAATAACTTTCCTAAGGCCACCCAGATAATAAGTGACAGAGCTGTGA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      40441 GCACAAAGAATGTAAATAACTTTCCTAAGGCCACCCAGATAATAAGTGACAGAGCTGTGA 40500
```

RESULT 2
US-09-925-065A-649342
; Sequence 649342, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:
;   APPLICANT: Wang, David G.
;   TITLE OF INVENTION: Identification and Mapping of Single
;   TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
;   FILE REFERENCE: 108827.135
;   CURRENT APPLICATION NUMBER: US/09/925,065A
;   CURRENT FILING DATE: 2001-08-08
;   PRIOR APPLICATION NUMBER: US 60/243,096
;   PRIOR FILING DATE: 2000-10-24
;   PRIOR APPLICATION NUMBER: US 60/252,147
;   PRIOR FILING DATE: 2000-11-20
;   PRIOR APPLICATION NUMBER: US 60/250,092
;   PRIOR FILING DATE: 2000-11-30
;   PRIOR APPLICATION NUMBER: US 60/261,766
;   PRIOR FILING DATE: 2001-01-16
;   PRIOR APPLICATION NUMBER: US 60/289,846
;   PRIOR FILING DATE: 2001-05-09
;   NUMBER OF SEQ ID NOS: 957086
;   SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 649342
;   LENGTH: 591
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-925-065A-649342

   Query Match   90.6%;   Score 543.6; DB 4;   Length 591;
   Best Local Similarity   99.8%;   Pred. No. 8.3e-82;
   Matches 543; Conservative   1; Mismatches   0; Indels   0; Gaps   0;

Qy   1 TACAATAGACCCTGCTTCTTTCAACGCTTCAACTGCTTTGCTATGTGTTACATCACGAAC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   48 TACAATAGACCCTGCTTCTTTCAACGCTTCAACTGCTTTGCTATGTGTTACATCACGAAC 107

Qy   61 ATCTACTTCATTTACTTGTAATATACAGTCATTGACCCTGAGAAAACAATTAAATATTAA 120
              |||||||||||||||:||||||||||||||||||||||||||||||||||||||||||||
Db   108 ATCTACTTCATTTACTYGTAATATACAGTCATTGACCCTGAGAAAACAATTAAATATTAA 167

Qy   121 AAATAAAAATTAAATATAATTTAAATATAAATTTTCCACTTTACCAATTTTTTGTTACTT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   168 AAATAAAAATTAAATATAATTTAAATATAAATTTTCCACTTTACCAATTTTTTGTTACTT 227

Qy   181 CTTTTTTAAGGTAAAGAGAATTATAAATAATTCTGGAGTAATTCCAGAAAACATAAATGA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   228 CTTTTTTAAGGTAAAGAGAATTATAAATAATTCTGGAGTAATTCCAGAAAACATAAATGA 287

Qy   241 AGAAAGTATATCAAAAACTAATATAAACAAATACAAACATTTCCCAAGGGCCAGCAAAAG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        288 AGAAAGTATATCAAAAACTAATATAAACAAATACAAACATTTCCCAAGGGCCAGCAAAAG 347

Qy        301 GAACAAAAGAAATAGTGAAAATAATAGATTATATAAAAATGTTAAATAATAATTACAGCT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        348 GAACAAAAGAAATAGTGAAAATAATAGATTATATAAAAATGTTAAATAATAATTACAGCT 407

Qy        361 ACCATTTGCATACTTCCTGCATGCTTATTCTGTGTAGGCATTAAGACACACTTTATAAAA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        408 ACCATTTGCATACTTCCTGCATGCTTATTCTGTGTAGGCATTAAGACACACTTTATAAAA 467

Qy        421 ATAGCAAACATTTATTTAGCACTAACCACATGCCAGGCACTTTCTTGGTATTTTAACCCT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        468 ATAGCAAACATTTATTTAGCACTAACCACATGCCAGGCACTTTCTTGGTATTTTAACCCT 527

Qy        481 CATGACACCTGTAAGCTTAATATATATTTTAATCCCTATTTCACAGATGGAGAAACTGAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        528 CATGACACCTGTAAGCTTAATATATATTTTAATCCCTATTTCACAGATGGAGAAACTGAG 587

Qy        541 GCAC 544
              ||||
Db        588 GCAC 591
```

RESULT 3
US-09-925-065A-649342
; Sequence 649342, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 649342
;   LENGTH: 591
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-925-065A-649342

  Query Match   90.6%; Score 543.6; DB 5; Length 591;
  Best Local Similarity   99.8%; Pred. No. 8.3e-82;
  Matches 543; Conservative   1; Mismatches   0; Indels   0; Gaps   0;

Qy   1 TACAATAGACCCTGCTTCTTTCAACGCTTCAACTGCTTTGCTATGTGTTACATCACGAAC 60
```

```
                     |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       48 TACAATAGACCCTGCTTCTTTCAACGCTTCAACTGCTTTGCTATGTGTTACATCACGAAC 107

Qy       61 ATCTACTTCATTTACTTGTAATATACAGTCATTGACCCTGAGAAAACAATTAAATATTAA 120
                     |||||||||||||||:|||||||||||||||||||||||||||||||||||||||||
Db      108 ATCTACTTCATTTACTYGTAATATACAGTCATTGACCCTGAGAAAACAATTAAATATTAA 167

Qy      121 AAATAAAAATTAAATATAATTTAAATATAAATTTTCCACTTTACCAATTTTTTGTTACTT 180
                     |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      168 AAATAAAAATTAAATATAATTTAAATATAAATTTTCCACTTTACCAATTTTTTGTTACTT 227

Qy      181 CTTTTTTAAGGTAAAGAGAATTATAAATAATTCTGGAGTAATTCCAGAAAACATAAATGA 240
                     |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      228 CTTTTTTAAGGTAAAGAGAATTATAAATAATTCTGGAGTAATTCCAGAAAACATAAATGA 287

Qy      241 AGAAAGTATATCAAAAACTAATATAAACAAATACAAACATTTCCCAAGGGCCAGCAAAAG 300
                     |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      288 AGAAAGTATATCAAAAACTAATATAAACAAATACAAACATTTCCCAAGGGCCAGCAAAAG 347

Qy      301 GAACAAAAGAAATAGTGAAAATAATAGATTATATAAAAATGTTAAATAATAATTACAGCT 360
                     |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      348 GAACAAAAGAAATAGTGAAAATAATAGATTATATAAAAATGTTAAATAATAATTACAGCT 407

Qy      361 ACCATTTGCATACTTCCTGCATGCTTATTCTGTGTAGGCATTAAGACACACTTTATAAAA 420
                     |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      408 ACCATTTGCATACTTCCTGCATGCTTATTCTGTGTAGGCATTAAGACACACTTTATAAAA 467

Qy      421 ATAGCAAACATTTATTTAGCACTAACCACATGCCAGGCACTTTCTTGGTATTTTAACCCT 480
                     |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      468 ATAGCAAACATTTATTTAGCACTAACCACATGCCAGGCACTTTCTTGGTATTTTAACCCT 527

Qy      481 CATGACACCTGTAAGCTTAATATATATTTTAATCCCTATTTCACAGATGGAGAAACTGAG 540
                     |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      528 CATGACACCTGTAAGCTTAATATATATTTTAATCCCTATTTCACAGATGGAGAAACTGAG 587

Qy      541 GCAC 544
                     ||||
Db      588 GCAC 591
```

RESULT 4
US-10-840-060-266/c
; Sequence 266, Application US/10840060
; Publication No: US20050227243A1
; GENERAL INFORMATION:
; APPLICANT: Cyclacel Limited
; APPLICANT: Deak, Peter
; APPLICANT: Frenz, Lisa
; APPLICANT: Glover, David
; APPLICANT: Midgley, Carol
; TITLE OF INVENTION: Cell Cycle Progression Proteins
; FILE REFERENCE: 10069/2012
; CURRENT APPLICATION NUMBER: US/10/840,060
; CURRENT FILING DATE: 2004-05-05
; PRIOR APPLICATION NUMBER: PCT/GB02/04780
; PRIOR FILING DATE: 2002-10-23
; PRIOR APPLICATION NUMBER: GB 0126506.5

```
  Query Match              16.3%;  Score 98;  DB 10;  Length 3046;
  Best Local Similarity    100.0%;  Pred. No. 1e-06;
  Matches   98; Conservative    0;  Mismatches    0;  Indels    0; Gaps    0;

Qy         1 TACAATAGACCCTGCTTCTTTCAACGCTTCAACTGCTTTGCTATGTGTTACATCACGAAC 60
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1097 TACAATAGACCCTGCTTCTTTCAACGCTTCAACTGCTTTGCTATGTGTTACATCACGAAC 1038

Qy        61 ATCTACTTCATTTACTTGTAATATACAGTCATTGACCC 98
             ||||||||||||||||||||||||||||||||||||||
Db      1037 ATCTACTTCATTTACTTGTAATATACAGTCATTGACCC 1000
```

RESULT 5
US-10-722-939-5/c

```
  Query Match              16.3%;  Score 97.6;  DB 10;  Length 3046;
  Best Local Similarity    99.0%;  Pred. No. 1.2e-06;
  Matches   97; Conservative    1;  Mismatches    0;  Indels    0; Gaps    0;
```

```
Qy        1 TACAATAGACCCTGCTTCTTTCAACGCTTCAACTGCTTTGCTATGTGTTACATCACGAAC 60
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1097 TACAATAGACCCTGCTTCTTTCAACGCTTCAACTGCTTTGCTATGTGTTACATCACGAAC 1038

Qy       61 ATCTACTTCATTTACTTGTAATATACAGTCATTGACCC 98
            ||||||||||||||||:|||||||||||||||||||||
Db     1037 ATCTACTTCATTTACTYGTAATATACAGTCATTGACCC 1000
```

RESULT 6
US-10-106-698-1545/c
; Sequence 1545, Application US/10106698
; Publication No. US20030109690A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and
Polypeptides
; FILE REFERENCE: PA005P1
; CURRENT APPLICATION NUMBER: US/10/106,698
; CURRENT FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: PCT/US00/26524
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US 60/157,137
; PRIOR FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: US 60/163,280
; PRIOR FILING DATE: 1999-11-03
; NUMBER OF SEQ ID NOS: 8564
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 1545
; LENGTH: 1152
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (6)..(6)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (17)..(17)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (64)..(64)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (1126)..(1126)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (1147)..(1147)
; OTHER INFORMATION: n equals a,t,g, or c
US-10-106-698-1545

   Query Match   16.0%; Score 96; DB 6; Length 1152;
   Best Local Similarity 98.0%; Pred. No. 1.7e-06;
   Matches   96; Conservative   1; Mismatches   1; Indels   0; Gaps   0;

```
Qy        1 TACAATAGACCCTGCTTCTTTCAACGCTTCAACTGCTTTGCTATGTGTTACATCACGAAC 60
            ||||:|||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        870 TACARTAGACCCTGCTTCTTTCAACGCTTCAACTGCTTTGCTATGTGTTACATCACGAAC 811

Qy         61 ATCTACTTCATTTACTTGTAATATACAGTCATTGACCC 98
              |||||||||||||||||| ||||||||||||||||||||
Db        810 ATCTACTTCATTTACTCGTAATATACAGTCATTGACCC 773
```

RESULT 7
US-09-960-352-11234/c
; Sequence 11234, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
;   APPLICANT: Warren, Wesley C.
;   APPLICANT: Tao, Nengbing
;   APPLICANT: Byatt, John C.
;   APPLICANT: Mathialagan, Nagappan
;   TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH
LACTATION AND
;   TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION
;   FILE REFERENCE: 16511.006/37-21(10298)C
;   CURRENT APPLICATION NUMBER: US/09/960,352
;   CURRENT FILING DATE: 2001-09-24
;   NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 11234
;   LENGTH: 419
;   TYPE: DNA
;   ORGANISM: Bos taurus
;   OTHER INFORMATION: Clone ID: 48-LIB3058-052-Q1-K1-D8
US-09-960-352-11234

     Query Match   13.8%; Score 82.6; DB 3; Length 419;
     Best Local Similarity   58.2%; Pred. No. 0.00022;
     Matches 145; Conservative   0; Mismatches 104; Indels   0; Gaps   0;

```
Qy         89 TCATTGACCCTGAGAAAACAATTAAATATTAAAAATAAAAATTAAATATAATTTAAATAT 148
              | ||| |   | | |||| || | ||| |||| |      ||| || |||| || |||
Db        274 TAATTTAAATTTAAAAAAATATAATATAAAAAAATTTTTTTTTATATTTAATAAAATTAT 215

Qy        149 AAATTTTCCACTTTACCAATTTTTTGTTACTTCTTTTTTTAAGGTAAAGAGAATTATAAAT 208
              |||||||     |||    ||||||||| |  || |||||| ||   ||| | || | |||
Db        214 AAATTTTAACTTTTTTAAAATTTTTTTATTATTATTTTTAAAAAAAAAAAAAAAAAAAAAA 155

Qy        209 AATTCTGGAGTAATTCCAGAAAACATAAATGAAGAAAGTATATCAAAAACTAATATAAAC 268
              ||      | || | | ||||| | ||| || ||| | | ||||| || | |||
Db        154 AAAAAAAAAAAAAATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 95

Qy        269 AAATACAAACATTTCCCAAGGGCCAGCAAAAGGAACAAAAGAAATAGTGAAAATAATAGA 328
              ||| | ||| |        ||     | |||| || |||| ||| |    |||| || | |
Db         94 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 35

Qy        329 TTATATAAA 337
              | | |||
Db         34 AAAAAAAAA 26
```

RESULT 8
US-10-473-126-386/c

```
    Query Match            12.8%;  Score 76.8;  DB 9;  Length 8056;
    Best Local Similarity  47.5%;  Pred. No. 0.0047;
    Matches  261; Conservative   0;  Mismatches 287;  Indels   2;  Gaps    1;

Qy        16 TTCTTTCAACGCTTCAACTGCTTTGCTATGTGTTACATCACGAACATCTACTTCATTTAC 75
             || ||| ||   || |  |   || || | || |        | || |||||
Db      2582 TTTTTTAAAAATTTTATTTAAATTTTTAAATTTTTTAATTTATTTTATAATTTAATTTAT 2523

Qy        76 TTGTAATATACAGTCATTGACCCTGAGAAAACAATTAAATATTAAAAATAAAAATTAAAT 135
             || |  |      ||||| |   | |   | |   | |||| | | || ||||||   |
Db      2522 TTTTTTTTATTTTTCATTAATATTTTAATTATTTTAAAACAAAATAATAAAAAATATATA 2463

Qy       136 ATAATTTAAATATAAATTTTCCACTTTACCAATTTTTTGTTACTTCTTTTTTAAGGTAAA 195
             | |   | | |  ||| ||    |||| ||| |     | || |||||||| |    |
Db      2462 AATAAAAATAAAAAAAAATTAATTTTTAAAAATAAATAAATTATTTTTTTTTTAATTTTC 2403

Qy       196 GAGAATTATAAATAATTCTGGAGTAATTCCAGAAAACATAAATGAAGAAAGTATATCAAA 255
               | || || || || | |   | | |||    |||| | |||  ||  | | |    ||
Db      2402 AAAAAATAATAAAATTTAATAAATTATTTATAAAAATAAAAAATTATATTTAAAAATTAA 2343

Qy       256 AACTAATATAAACAAATACAAACATTTCCCAAGGGCCAGCAAAAGGAACAAAAGAAATAG 315
             |   ||| ||  |||||||||| |||   ||      |||| || ||| |||| |
Db      2342 ATAAAATTTATTTAAATACAAAAATTAATAAATTATTTTAAAAATAAATAAATTAAAAAA 2283

Qy       316 TGAAAA--TAATAGATTATATAAAAATGTTAAATAATAATTACAGCTACCATTTGCATAC 373
             | |||| | | |  ||| | |||| |||| |||||| | | |    | | || ||
Db      2282 TTAAAATTTTTTTAATTTATTTAAATATTAAAAAATAAATTTTTATTAAAATTAAATTT 2223

Qy       374 TTCCTGCATGCTTATTCTGTGTAGGCATTAAGACACACTTTATAAAAATAGCAAACATTT 433
             || |  | |  || | |  || || |  | ||  || ||||||| || |       |||
Db      2222 TTTTTTTTTTTATTTTTTTTTTCATTTTTTTAAATTTTTAATAAAATTATAATTTTTTT 2163

Qy       434 ATTTAGCACTAACCACATGCCAGGCACTTTCTTGGTATTTTAACCCTCATGACACCTGTA 493
             |||   | |   | |   | |  ||||| || | |  | |   | ||
Db      2162 TTTTTTTATTTTTTATTTTTAATAAAATAAAAAATTATTAAAATTATATTAAATAATTTA 2103

Qy       494 AGCTTAATATATATTTTAATCCCTATTTCACAGATGGAGAAACTGAGGCACAAAGAATGT 553
             ||| ||| ||| || |   || |    || |      | |    | | ||| |||| ||| |
```

```
Db      2102 TAATTATTATTTATAATATTTAAAATATTTTTAATAAAATTTTTTAAAAATAAAAAATAT 2043

Qy       554 AAATAACTTT 563
              |  || |  |
Db      2042 TTAAAAATAT 2033


RESULT 9
US-10-425-115-7873/c
; Sequence 7873, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
;   APPLICANT: La Rosa, Thomas J.
;   APPLICANT:  Kovalic, David K.
;   APPLICANT:  Zhou, Yihua
;   APPLICANT:  Cao, Yongwei
;   TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated
With
;   TITLE OF INVENTION:  Plants
;   FILE REFERENCE: 38-21(53222)B
;   CURRENT APPLICATION NUMBER: US/10/425,115
;   CURRENT FILING DATE:  2003-04-28
;   NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 7873
;   LENGTH: 542
;   TYPE: DNA
;   ORGANISM: Zea mays
;   FEATURE:
;   OTHER INFORMATION: Clone ID: MRT4577_107175C.1
US-10-425-115-7873

   Query Match              12.8%;   Score 76.6;  DB 9;  Length 542;
   Best Local Similarity    56.6%;   Pred. No. 0.0024;
   Matches  142;  Conservative    0;  Mismatches  109;  Indels    0;  Gaps    0;

Qy       105 AACAATTAAATATTAAAAATAAAAATTAAATATAATTTAAATATAAATTTTCCACTTTAC 164
              || |||||  | |  ||||| ||||| |||||| ||    | || ||||     |||
Db       344 AAAAATTTTAAAAAAAAAAAAAAAAAAAAAAAATATTTTTATGTTTTAGTTTTTTTTTTTTT 285

Qy       165 CAATTTTTTGTTACTTCTTTTTTTAAGGTAAAGAGAATTATAAATAATTCTGGAGTAATTC 224
                ||      |||  || ||||||||     ||| | ||  | ||| ||      |  ||
Db       284 AAAAAAAACGTTTTTTTTTTTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 225

Qy       225 CAGAAAACATAAATGAAGAAAGTATATCAAAAACTAATATAAACAAATACAAACATTTCC 284
              | |||| | |||  || |||  | |  ||||| || | ||| ||| | |||| |
Db       224 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 165

Qy       285 CAAGGGCCAGCAAAAGGAACAAAAGAAATAGTGAAAATAATAGATTATATAAAAATGTTA 344
              ||      |  ||||  || ||||| ||| |    |||| || | |  | | |||||  |
Db       164 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 105

Qy       345 AATAATAATTA 355
              || || || ||
Db       104 AAAAAAAAATA 94


RESULT 10
```

US-10-486-706-306/c
; Sequence 306, Application US/10486706
; Publication No. US20050071088A1
; GENERAL INFORMATION:
; APPLICANT: LANDFIELD, PHILIP W.
; APPLICANT: BLALOCK, ERIC M.
; APPLICANT: CHEN, KUEY-CHU
; APPLICANT: FOSTER, THOMAS C.
; TITLE OF INVENTION: GENE EXPRESSION PROFILE BIOMARKERS AND THERAPEUTIC
TARGETS FOR
; TITLE OF INVENTION: BRAIN AGING AND AGE-RELATED COGNITIVE IMPAIRMENT
; FILE REFERENCE: 50229-426
; CURRENT APPLICATION NUMBER: US/10/486,706
; CURRENT FILING DATE: 2004-02-13
; PRIOR APPLICATION NUMBER: PCT/US02/25607
; PRIOR FILING DATE: 2002-08-13
; PRIOR APPLICATION NUMBER: US 60/311,343
; PRIOR FILING DATE: 2001-08-13
; NUMBER OF SEQ ID NOS: 461
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 306
;   LENGTH: 3256
;   TYPE: DNA
;   ORGANISM: Rattus norvegicus
US-10-486-706-306

   Query Match   12.7%; Score 76.2; DB 10; Length 3256;
   Best Local Similarity   86.6%; Pred. No. 0.0047;
   Matches   84; Conservative   0; Mismatches   13; Indels   0; Gaps   0;

Qy   2 ACAATAGACCCTGCTTCTTTCAACGCTTCAACTGCTTTGCTATGTGTTACATCACGAACA 61
             || || || || ||||||||||| |||||||||||||||| || |||||||||||||||
Db   1359 ACTATCGATCCAGCTTCTTTCAATGCTTCAACTGCTTTGCTGTGGGGTTACATCACGAACA 1300

Qy   62 TCTACTTCATTTACTTGTAATATACAGTCATTGACCC 98
             ||| ||||||||||| | |||||||||||| ||||
Db   1299 TCTGCTTCATTTACTCTCAGTATACAGTCATTTACCC 1263

RESULT 11
US-10-980-571-14/c
; Sequence 14, Application US/10980571
; Publication No. US20050138675A1
; GENERAL INFORMATION:
; APPLICANT: Pfizer, Inc.
; TITLE OF INVENTION: Method For Determining Cardiotoxicity
; FILE REFERENCE: PC26200A
; CURRENT APPLICATION NUMBER: US/10/980,571
; CURRENT FILING DATE: 2004-11-02
; PRIOR APPLICATION NUMBER: 60/516,774
; PRIOR FILING DATE: 2003-11-03
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 14
;   LENGTH: 3256
;   TYPE: DNA
;   ORGANISM: Rattus norvegicus

```
Query Match              12.7%;  Score 76.2;  DB 10;  Length 3256;
Best Local Similarity    86.6%;  Pred. No. 0.0047;
Matches   84;  Conservative    0;  Mismatches  13;  Indels    0;  Gaps    0;

Qy        2 ACAATAGACCCTGCTTCTTTCAACGCTTCAACTGCTTTGCTATGTGTTACATCACGAACA 61
            || || || ||  ||||||||||| |||||||||||||||| || ||||||||||||||
Db     1359 ACTATCGATCCAGCTTCTTTCAATGCTTCAACTGCTTTGCTGTGGGTTACATCACGAACA 1300

Qy       62 TCTACTTCATTTACTTGTAATATACAGTCATTGACCC 98
            ||| |||||||||||    | ||||||||||||| ||||
Db     1299 TCTGCTTCATTTACTCTCAGTATACAGTCATTTACCC 1263


RESULT 12
US-10-473-126-240/c
; Sequence 240, Application US/10473126
; Publication No. US20040234973A1
; GENERAL INFORMATION:
;   APPLICANT: Epigenomics AG
;   TITLE OF INVENTION: Methods and nucleic acids for the analysis of
hematopoietic cell
;   TITLE OF INVENTION:  proliferative disorders
;   FILE REFERENCE:
;   CURRENT APPLICATION NUMBER: US/10/473,126
;   CURRENT FILING DATE:  2003-09-26
;   NUMBER OF SEQ ID NOS: 1258
; SEQ ID NO 240
;   LENGTH: 8056
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-473-126-240

Query Match              12.5%;  Score 75.2;  DB 9;  Length 8056;
Best Local Similarity    47.3%;  Pred. No. 0.0088;
Matches  260;  Conservative    0;  Mismatches 288;  Indels    2;  Gaps    1;

Qy       16 TTCTTTCAACGCTTCAACTGCTTTGCTATGTGTTACATCACGAACATCTACTTCATTTAC 75
            || ||| ||   || |  | || || | ||         | || |||||
Db     2582 TTTTTTAAAAATTTTATTTAAATTTTTAAATTTTTTAATTTATTTTATAATTTAATTTAT 2523

Qy       76 TTGTAATATACAGTCATTGACCCTGAGAAAACAATTAAATATTAAAAATAAAAATTAAAT 135
            || |  |      || ||| |  |  | | | |  | |||   | || ||||||  |
Db     2522 TTTTTTTTATTTTTCGTTAATATTTTAATTATTTTAAAACGAAATAATAAAAAATATATA 2463

Qy      136 ATAATTTAAATATAAATTTTCCACTTTACCAATTTTTTGTTACTTCTTTTTTAAGGTAAA 195
            |  |    | | | ||| ||        |||| |||  |   |  || |||||||  |  |
Db     2462 AATAAAAATAAAAAAAAATTAATTTTTAAAAATAAATAAATTATTTTTTTTTTAATTTTC 2403

Qy      196 GAGAATTATAAATAATTCTGGAGTAATTCCAGAAAACATAAATGAAGAAAGTATATCAAA 255
            || || ||  || | ||     | | |||     |||| | |||  | |   | |   ||
Db     2402 GAAAAATAATAAAATTTAATAAATTATTTATAAAAATAAAAAATTATATTTAAAAATTAA 2343

Qy      256 AACTAATATAAACAAATACAAACATTTCCCAAGGGCCAGCAAAAGGAACAAAAGAAATAG 315
```

OM nucleic - nucleic search, using sw model

Run on: July 15, 2006, 18:06:17 ; Search time 4521 Seconds
                                           (without alignments)
                                           7421.278 Million cell updates/sec

Title: US-10-722-939-1-T39977_39901-40500
Perfect score: 600
Sequence: 1 tacaatagaccctgcttctt..........aataagtgacagagctgtga 600

Scoring table: IDENTITY_NUC
                 Gapop 10.0 , Gapext 1.0

Searched: 48236798 seqs, 27959665780 residues

Total number of hits satisfying chosen parameters: 96473596

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 100 summaries

Database :   EST:*
                 1: gb_est1:*
                 2: gb_est3:*
                 3: gb_est4:*
                 4: gb_est5:*
                 5: gb_est6:*
                 6: gb_htc:*
                 7: gb_est2:*
                 8: gb_est7:*
                 9: gb_est8:*
                 10: gb_est9:*
                 11: gb_gss1:*
                 12: gb_gss2:*
                 13: gb_gss3:*
                 14: gb_gss4:*

          Pred. No. is the number of results predicted by chance to have a
          score greater than or equal to the score of the result being printed,
          and is derived by analysis of the total score distribution.

                                    SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
c 1	138.4	23.1	740	2	BG567499	BG567499 602586233
2	136.8	22.8	906	12	CC504900	CC504900 CH240_345
3	96.8	16.1	462	10	H96495	H96495 yt98h02.r1

	4	96.4	16.1	361	1	AA281170	AA281170 zt01b11.r
c	5	96.4	16.1	362	9	DA501101	DA501101 DA501101
	6	96.4	16.1	482	1	AA831793	AA831793 oa59d10.s
	7	96.4	16.1	531	3	BU741256	BU741256 UI-E-EJ0-
c	8	96.4	16.1	532	3	BM727000	BM727000 UI-E-EJ0-
c	9	96.4	16.1	578	9	DA315109	DA315109 DA315109
c	10	96.4	16.1	580	9	DA262508	DA262508 DA262508
c	11	96.4	16.1	842	7	BF677732	BF677732 602085430
c	12	96.4	16.1	860	2	BG164149	BG164149 602343020
c	13	96.4	16.1	2675	14	AY420526	AY420526 Pan trogl
	14	94.8	15.8	447	1	AA480937	AA480937 aa28c06.r
c	15	94.8	15.8	582	3	BP229026	BP229026 BP229026
c	16	94.8	15.8	590	3	BP211156	BP211156 BP211156
c	17	92.8	15.5	601	8	CR762761	CR762761 DKFZp469I
c	18	90	15.0	653	9	DN267047	DN267047 LIB30320_
c	19	87.2	14.5	1309	7	BE420736	BE420736 HWM002.A0
	20	86.8	14.5	732	4	BX668084	BX668084 BX668084
	21	84.2	14.0	794	11	AZ530415	AZ530415 ENTBX01TR
	22	83.8	14.0	1297	10	DV778934	DV778934 Hw_Fat_9_
	23	83.2	13.9	1101	14	CNS00EVL	AL069706 Drosophil
c	24	82.8	13.8	867	14	CNS00CX5	AL060052 Drosophil
	25	82.6	13.8	866	5	CF289423	CF289423 AGENCOURT
c	26	82.4	13.7	1355	14	AG346348	AG346348 Mus muscu
c	27	81.8	13.6	1116	2	BM477099	BM477099 AGENCOURT
	28	81.4	13.6	464	9	CX571966	CX571966 TTE000181
c	29	81	13.5	729	8	CV486140	CV486140 AGENCOURT
c	30	80.4	13.4	1114	3	BQ422058	BQ422058 AGENCOURT
c	31	80	13.3	1203	14	CNS015WU	AL106008 Drosophil
c	32	79.6	13.3	993	12	CL113812	CL113812 ISB1-59A1
c	33	79.4	13.2	839	3	BQ151187	BQ151187 NF048A07L
	34	79.2	13.2	1101	14	CNS0183Y	AL108856 Drosophil
	35	79	13.2	983	1	AJ926697	AJ926697 AJ926697
c	36	79	13.2	1015	12	CL135318	CL135318 ISB1-106F
	37	78.6	13.1	996	14	CNS00FUH	AL071063 Drosophil
	38	78.6	13.1	1101	14	CNS003BD	AL064091 Drosophil
	39	78.4	13.1	759	14	CNS06QXV	AL411257 T7 end of
c	40	78.4	13.1	1007	10	DV050335	DV050335 DAY35S_08
c	41	78.2	13.0	1231	12	CG754046	CG754046 P049-2-D0
	42	77.8	13.0	1101	14	CNS00LT2	AL078714 Drosophil
c	43	77.8	13.0	1255	14	AG346372	AG346372 Mus muscu
c	44	77.6	12.9	1007	9	DN566957	DN566957 92914460
	45	77.6	12.9	1088	1	AJ928262	AJ928262 AJ928262
c	46	77.6	12.9	1283	13	CL641341	CL641341 CH213-9O1
	47	77.4	12.9	658	14	CNS03NK1	AL252154 Tetraodon
	48	77.4	12.9	661	10	DR722785	DR722785 AGENCOURT
c	49	77.4	12.9	764	5	CF289381	CF289381 AGENCOURT
c	50	77.2	12.9	811	1	AL514901	AL514901 AL514901
c	51	77	12.8	1249	14	AG347294	AG347294 Mus muscu
c	52	76.8	12.8	834	14	CNS00868	AL051779 Drosophil
c	53	76.8	12.8	1101	14	CNS017V2	AL108536 Drosophil
c	54	76.8	12.8	1150	14	AG365432	AG365432 Mus muscu
c	55	76.6	12.8	1335	14	AG290024	AG290024 Mus muscu
c	56	76.4	12.7	989	11	AZ549932	AZ549932 ENTFP96TF
c	57	76.4	12.7	1155	10	DW671774	DW671774 CNB367-C0
	58	76.2	12.7	341	1	AA849677	AA849677 EST192444
	59	76.2	12.7	410	1	AI144926	AI144926 UI-R-BT0-
	60	76.2	12.7	801	8	CV485384	CV485384 AGENCOURT

	61	76.2	12.7	826	8	CV483860	CV483860 AGENCOURT
c	62	76	12.7	1101	14	CNS016LI	AL106896 Drosophil
c	63	75.8	12.6	1101	14	CNS00EO7	AL069440 Drosophil
c	64	75.8	12.6	1269	14	AG386989	AG386989 Mus muscu
c	65	75.8	12.6	1880	14	AG430399	AG430399 Mus muscu
	66	75.6	12.6	527	4	CA726232	CA726232 wet1s.pk0
c	67	75.6	12.6	576	14	CNS035N7	AL228940 Tetraodon
c	68	75.6	12.6	745	1	AJ445424	AJ445424 AJ445424
c	69	75.6	12.6	990	14	CNS006OI	AL065624 Drosophil
c	70	75.6	12.6	1101	14	CNS002FG	AL062437 Drosophil
c	71	75.6	12.6	1101	14	CNS00ETW	AL069847 Drosophil
c	72	75.2	12.5	897	14	CNS07ABZ	AL436389 T7 end of
c	73	74.8	12.5	947	12	CL112501	CL112501 ISB1-57D2
c	74	74.8	12.5	1244	14	AG361248	AG361248 Mus muscu
	75	74.8	12.5	1260	13	CL491610	CL491610 SAIL_559_
	76	74.8	12.5	1522	12	CL128484	CL128484 ISB1-94I1
c	77	74.6	12.4	500	9	DN386720	DN386720 LIB3892-0
	78	74.6	12.4	607	8	CV490737	CV490737 AGENCOURT
c	79	74.6	12.4	891	14	CNS009JU	AL053767 Drosophil
	80	74.6	12.4	1058	12	CL077132	CL077132 CH216-143
	81	74.6	12.4	1101	14	CNS0039G	AL063921 Drosophil
c	82	74.6	12.4	1542	14	AG386981	AG386981 Mus muscu
	83	74.4	12.4	734	14	CNS010MP	AL099163 Drosophil
c	84	74.4	12.4	749	11	AQ324504	AQ324504 mgxb0018L
c	85	74.4	12.4	920	9	DN564190	DN564190 90895070
	86	74.4	12.4	945	14	CNS04D0K	AL285149 Tetraodon
c	87	74.4	12.4	946	13	CZ934132	CZ934132 251587 To
c	88	74.4	12.4	1027	14	CNS02T50	AL212733 Tetraodon
c	89	74.4	12.4	1029	14	CNS01ZGM	AL174271 Tetraodon
c	90	74.4	12.4	1260	14	AG320287	AG320287 Mus muscu
c	91	74.4	12.4	1295	12	CG757539	CG757539 P052-4-E0
c	92	74.2	12.4	928	14	CNS00DKY	AL071865 Drosophil
	93	74.2	12.4	934	13	CL506649	CL506649 SAIL_770_
c	94	74.2	12.4	1074	12	BZ696936	BZ696936 SP__Ba009
c	95	74.2	12.4	1225	14	CNS0161D	AL106171 Drosophil
c	96	74	12.3	1101	14	CNS0145U	AL103740 Drosophil
c	97	73.8	12.3	998	12	BZ695174	BZ695174 SP__Ba005
c	98	73.8	12.3	1025	14	CNS014J2	AL104216 Drosophil
c	99	73.8	12.3	1227	14	AG430010	AG430010 Mus muscu
	100	73.6	12.3	1184	14	CNS04P4P	AL300850 Tetraodon

ALIGNMENTS

RESULT 1
BG567499/c
LOCUS   BG567499   740 bp   mRNA   linear   EST 10-APR-2001
DEFINITION   602586233F1 NIH_MGC_76 Homo sapiens cDNA clone IMAGE:4715024 5',
                mRNA sequence.
ACCESSION   BG567499
VERSION   BG567499.1 GI:13575152
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
  ORGANISM   Homo sapiens
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;

                    Hominidae; Homo.
REFERENCE   1   (bases 1 to 740)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
             Email: cgapbs-r@mail.nih.gov
             Tissue Procurement: CLONETECH Laboratories, Inc.
              cDNA Library Preparation: CLONETECH Laboratories, Inc.
              cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
              DNA Sequencing by: Incyte Genomics, Inc.
              Clone distribution: MGC clone distribution information can be
             found through the I.M.A.G.E. Consortium/LLNL at:
             http://image.llnl.gov
             Plate: LLCM1561 row: a column: 09
             High quality sequence stop: 474.
FEATURES   Location/Qualifiers
     source   1. .740
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4715024"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /clone_lib="NIH_MGC_76"
                     /note="Organ: liver; Vector: pDNR-LIB (Clontech); Site_1:
                     SfiI (ggccgcctcggcc); Site_2: SfiI (ggccattatggcc); 5' and
                     3' adaptors were used in cloning as follows: 5' adaptor
                     sequence: 5'-CACGGCCATTATGGCC-3' and 3' adaptor sequence:
                     5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3' (where B = A,
                     C, or G and N = A, C, G, or T). Average insert size 1.85
                     kb (range 1.0-4.0 kb). 15/15 colonies contained inserts
                     by PCR. This library was enriched for full-length clones
                     and was constructed by Clontech Laboratories (Palo Alto,
                     CA). Note: this is a NIH_MGC Library."
ORIGIN

   Query Match   23.1%; Score 138.4; DB 2; Length 740;
   Best Local Similarity   99.3%; Pred. No. 8.6e-14;
     Matches 139; Conservative   0; Mismatches   1; Indels   0; Gaps   0;

Qy   461 TTTCTTGGTATTTTAACCCTCATGACACCTGTAAGCTTAATATATATTTTAATCCCTATT 520
           ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   452 TTTTTTGGTATTTTAACCCTCATGACACCTGTAAGCTTAATATATATTTTAATCCCTATT 393

Qy   521 TCACAGATGGAGAAACTGAGGCACAAAGAATGTAAATAACTTTCCTAAGGCCACCCAGAT 580
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   392 TCACAGATGGAGAAACTGAGGCACAAAGAATGTAAATAACTTTCCTAAGGCCACCCAGAT 333

Qy   581 AATAAGTGACAGAGCTGTGA 600
           ||||||||||||||||||||
Db   332 AATAAGTGACAGAGCTGTGA 313

RESULT 2
CC504900
LOCUS   CC504900   906 bp   DNA   linear   GSS 17-JUN-2003
DEFINITION   CH240_345A2.T7 CHORI-240 Bos taurus genomic clone CH240_345A2,

```
                    genomic survey sequence.
ACCESSION    CC504900
VERSION      CC504900.1  GI:31823193
KEYWORDS     GSS.
SOURCE       Bos taurus (cattle)
  ORGANISM   Bos taurus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
             Pecora; Bovidae; Bovinae; Bos.
REFERENCE    1  (bases 1 to 906)
  AUTHORS    Holt,R., Stott,J., Yang,G., Barber,S., Smailus,D., Prabhu,A.-L.,
             Tsai,M., Cloutier,A., Lee,D., Girn,N., Olson,T., Mayo,M.,
             Butterfield,Y., Kirkpatrick,R., Liu,J., Guin,R., Chan,A., Chiu,R.,
             Mathewson,C., Wye,N., Masson,A., Brown-John,M., Jones,S.,
             Schein,J., Marra,M., de Jong,P., McWilliam,S., Barris,W.,
             Dalrymple,B.P. and Tellam,R.
  TITLE      Bovine BAC End Sequences from Library CHORI-240, PLATES 294 to 398
  JOURNAL    Unpublished (2003)
COMMENT      Other_GSSs: CH240_345A2.TARBAC13P2
             Contact: Rob Holt
             Sequencing
             The British Columbia Cancer Agency Genome Science Centre
             600 W. 10th Ave, Vancouver, British Columbia, Canada V5Z 4E6
             Tel: 604-877-6085
             Fax: 604-877-6276
             Email: rholt@bcgsc.ca
             Clones are derived from the bovine BAC library CHORI-240
             (http://www.chori.org/bacpac/bovine240.htm). For BAC library
             availability, please contact Pieter de Jong (pdejong@mail.cho.org).
             Clones may be purchased from BACPAC Resources
             (http://www.chori.org/bacpac/ordering_information.htm). This work
             was undertaken as part of the International Bovine BAC Mapping
             Consortium (IBBMC) by CSIRO Livestock Industries, Australia and the
             British Columbia Genome Sciences Centre, Canada.
             Plate: 345  row: A  column: 2
             Seq primer: T7
             Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .906
                     /organism="Bos taurus"
                     /mol_type="genomic DNA"
                     /strain="breed: Hereford"
                     /db_xref="taxon:9913"
                     /clone="CH240_345A2"
                     /sex="Male"
                     /cell_type="Blood"
                     /clone_lib="CHORI-240"
                     /note="Vector: pTARBAC1.3; Site_1: MboI; Site_2: MboI;
                     Hereford bull L1 Domino 99375; CHORI-240 Bovine BAC
                     library (Male) produced by Pieter de Jong"
ORIGIN

  Query Match              22.8%;  Score 136.8;  DB 12;  Length 906;
  Best Local Similarity    63.6%;  Pred. No. 1.6e-13;
  Matches  347; Conservative    0; Mismatches 147; Indels   52; Gaps    7;

Qy        55 ACGAACATCTACTTCATTTACTTGTAATATACAGTCATTGACCCTGAGAAAACAATTAAA 114
```

```
          ||||||||||| |||||||||| | || |||||||||||||| || || ||| |||||||
Db      1 ACGAACATCTATTTCATTTACTCGCAAAATACAGTCATTGACTCTAAGGAAA-AATTAAA 59

Qy    115 TATTAAAAATAAAAATTAAATATAATTTAAATATAAATTTTCCACTTTACCAATTTTTTG 174
            || |||| ||| |||||    || || ||||| ||||| ||||||| |||||||||
Db     60 AATAAAAATGAAATATTAA----AAATTTAATATCGATTTTTCACTTTGCCAATTTTTAT 115

Qy    175 TTACTTCTTTTTTAAGGTAAAGAGAATTATAAATAATTCTGGAGTAATTCCAGAAAACAT 234
            || |    |||||| | | ||| |  |  ||   | |        | |||  | ||
Db    116 TTTCACATTTTTAAGGTTAACTACTACGATTTTTTCTGAGAAGAAACTTCATGTTAA--- 172

Qy    235 AAATGAAGAAAGTATATCAAAAACTAATATAAACAAATACAAACATTTCCCAAGGGCCAG 294
              |    | ||||||||| ||| ||||    |||||||||||||||| |||||||||
Db    173 -----ATTTCAATATATCAAATACTTATAT----AAATACAAACATTTCTCAAGGGCCAT 223

Qy    295 CAAAAGGAACAAAAGAAATAGTGAAAATAATAGATTATATAAAAATGTTAAATAATAATT 354
            |||| | || |     ||||     |||||    ||| || ||||||| ||||||| |||||
Db    224 CAAAGGAAATAGTGCAAAT------AATAGATTATTTTACAAAAATATTAAATGATAATG 277

Qy    355 ACAGCTACCATTTGCATACTTCCTGCATGCTTATTCTGTGTAGGCATTAAGACACACTTT 414
            ||||||                            ||| |||| |     | | ||||||
Db    278 ACAGCT----------------------------GTCTATGTACAGAAATTAAAATACTTT 310

Qy    415 ATAAAAATAGCAAACATTTATTTAGCACTAACCACATGCCAGGCACTTTCTTGGTATTTT 474
            | |||||| || ||| | ||||||||| ||||| |||| || || |||| ||      |||
Db    311 AGAAAACAGAAAATACTTATTTAGTACTAATCACAAGCTACAAACTTCTTTTAGTGTTT 370

Qy    475 AACCCTCATGACACCTGTAAGCTTAATATATATTTTAATCCCTATTTCACAGATGGAGAA 534
            | ||     |       |        || ||| | |||||||||| ||| ||||| ||
Db    371 TAACCCTTACAACAACCTGTAAGGTGGATCTATCTATATCCCTATTTTACACATGGAAAA 430

Qy    535 ACTGAGGCACAAAGAATGTAAATAACTTTCCTAAGGCCACCCAGATAATAAGTGACAGAG 594
            ||||||||||||||| | ||||||||| |||| || |||||||||||||| |     ||||
Db    431 ACTGAGGCACAAAG--TTTAAATAACATTCCCAAAGCCACCCAGATAATGACCAGTAGAG 488

Qy    595 CTGTGA 600
            | ||||
Db    489 CAGTGA 494


RESULT 3
H96495
LOCUS       H96495                   462 bp    mRNA      linear   EST 25-NOV-1996
DEFINITION  yt98h02.r1 Soares_pineal_gland_N3HPG Homo sapiens cDNA clone
            IMAGE:232371 5' similar to SP:A45436 A45436
            SAP90=SYNAPSE-ASSOCIATED PROTEIN - ;, mRNA sequence.
ACCESSION   H96495
VERSION     H96495.1  GI:1109963
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 462)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
```

                    Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
                    Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
                    Wilson,R.
  TITLE   The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
  COMMENT   Contact: Wilson RK
                    Washington University School of Medicine
                    4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
                    Tel: 314 286 1800
                    Fax: 314 286 1810
                    Email: est@watson.wustl.edu
                    High quality sequence stops: 333
                    Source: IMAGE Consortium, LLNL
                    This clone is available royalty-free through LLNL ; contact the
                    IMAGE Consortium (info@image.llnl.gov) for further information.
                    Possible reversed clone: similarity on wrong strand
                    Insert Length: 2241   Std Error: 0.00
                    Seq primer: M13RP1.
FEATURES   Location/Qualifiers
     source   1. .462
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="GDB:3862362"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:232371"
                     /lab_host="DH10B (ampicillin resistant)"
                     /clone_lib="Soares_pineal_gland_N3HPG"
                     /note="Organ: pineal gland; Vector: pT7T3D (Pharmacia)
                     with a modified polylinker; Site_1: Not I; Site_2: Eco RI;
                     1st strand cDNA was primed with a Not I - oligo(dT) primer
                     [5' TGTTACCAATCTGAAGTGGGAGCGGCCGCCGTTTTTTTTTTTTTTTTTTTT
                     3'], double-stranded cDNA was size selected, ligated to
                     Eco RI adapters (Pharmacia), digested with Not I and
                     cloned into the Not I and Eco RI sites of a modified pT7T3
                     vector (Pharmacia). Library constructed by Bento Soares
                     and M.Fatima Bonaldo. "
ORIGIN

   Query Match   16.1%;   Score 96.8;   DB 10;   Length 462;
   Best Local Similarity   98.0%;   Pred. No. 1.1e-06;
   Matches   98; Conservative   0; Mismatches   2; Indels   0; Gaps   0;

Qy   1 TACAATAGACCCTGCTTCTTTCAACGCTTCAACTGCTTTGCTATGTGTTACATCACGAAC 60
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   199 TACAATAGACCCTGCTTCTTTCAACGCTTCAACTGCTTTGCTATGTGTTACATCACGAAC 258

Qy   61 ATCTACTTCATTTACTTGTAATATACAGTCATTGACCCTG 100
             |||||||||||||||||| |||||||||||||||||||| |
Db   259 ATCTACTTCATTTACTCGTAATATACAGTCATTGACCCGG 298

RESULT 4
AA281170
LOCUS   AA281170   361 bp   mRNA   linear   EST 14-AUG-1997
DEFINITION zt01b11.r1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:711837 5'
                    similar to TR:G558438 G558438 HOMOLOG OF DROSOPHILA DISCS LARGE
                    PROTEIN, ISOFORM 1. ;, mRNA sequence.

```
ACCESSION    AA281170
VERSION      AA281170.1  GI:1923851
KEYWORDS     EST.
SOURCE       Homo sapiens (human)
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
             Hominidae; Homo.
REFERENCE    1  (bases 1 to 361)
  AUTHORS    NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE      National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
             Tumor Gene Index
  JOURNAL    Unpublished (1997)
COMMENT      Contact: Robert Strausberg, Ph.D.
             Email: cgapbs-r@mail.nih.gov
             This clone is available royalty-free through LLNL ; contact the
             IMAGE Consortium (info@image.llnl.gov) for further information.
             Possible reversed clone: similarity on wrong strand
             Insert Length: 1961    Std Error: 0.00
             Seq primer: -28m13 rev2 ET from Amersham
             High quality sequence stop: 360.
FEATURES             Location/Qualifiers
     source          1. .361
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:711837"
                     /tissue_type="germinal center B cell"
                     /lab_host="DH10B"
                     /clone_lib="NCI_CGAP_GCB1"
                     /note="Vector: pT7T3D-PacI; Site_1: Not I; Site_2: Eco RI;
                     1st strand cDNA was prepared from human tonsillar cells
                     enriched for germinal center B cells by flow sorting
                     (CD20+, IgD-), provided by Dr. Louis M. Staudt (NCI), Dr.
                     David Allman (NCI) and Dr. Gerald Marti (CBER).  cDNA
                     synthesis was primed with a Not I - oligo(dT) primer
                     [5'-TGTTACCAATCTGAAGTGGGAGCGGCCGCCTCATTTTTTTTTTTTTTTTTT-3'
                     ]. Double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization, and was
                     constructed by Bento Soares and M. Fatima Bonaldo."
ORIGIN
```

```
   Query Match            16.1%;  Score 96.4;  DB 1;  Length 361;
   Best Local Similarity  99.0%;  Pred. No. 1.3e-06;
     Matches   97; Conservative    0; Mismatches    1; Indels    0; Gaps    0;


Qy       1 TACAATAGACCCTGCTTCTTTCAACGCTTCAACTGCTTTGCTATGTGTTACATCACGAAC 60
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      34 TACAATAGACCCTGCTTCTTTCAACGCTTCAACTGCTTTGCTATGTGTTACATCACGAAC 93


Qy      61 ATCTACTTCATTTACTTGTAATATACAGTCATTGACCC 98
           ||||||||||||||||| |||||||||||||||||||||
Db      94 ATCTACTTCATTTACTCGTAATATACAGTCATTGACCC 131
```

RESULT 5
DA501101/c
LOCUS   DA501101   362 bp   mRNA   linear   EST 08-NOV-2005
DEFINITION   DA501101 FCBBF3 Homo sapiens cDNA clone FCBBF3015237 5', mRNA
                sequence.
ACCESSION   DA501101
VERSION   DA501101.1 GI:81184123
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
  ORGANISM   Homo sapiens
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
                Hominidae; Homo.
REFERENCE   1 (bases 1 to 362)
  AUTHORS   Kimura,K., Wakamatsu,A., Suzuki,Y., Ota,T., Nishikawa,T.,
                Yamashita,R., Yamamoto,J., Sekine,M., Tsuritani,K., Wakaguri,H.,
                Ishii,S., Sugiyama,T., Saito,K., Isono,Y., Irie,R., Kushida,N.,
                Yoneyama,T., Otsuka,R., Kanda,K., Yokoi,T., Kondo,H., Wagatsuma,M.,
                Murakawa,K., Ishida,S., Ishibashi,T., Takahashi-Fujii,A.,
                Tanase,T., Nagai,K., Kikuchi,H., Nakai,K., Isogai,T. and Sugano,S.
  TITLE   Diversification of Transcriptional Modulation: Large-scale
                Identification and Characterization of Putative Alternative
                Promoters of Human Genes
  JOURNAL   Genome Res. 16 (1), 55-65 (2006)
  PUBMED   16344560
COMMENT   Contact: Takao Isogai
                FLJ Project (HRI Team)
                Helix Research Institute
                2-6-7 Kazusa-Kamatari, Kisarazu, Chiba, 292-0818, Japan
                Tel: 81-438-52-3975
                Fax: 81-438-52-3986
                Email: flj-cdna@nifty.com
                NEDO human cDNA project (New Energy and Industrial Technology
                Developmental Organization, Japan); cDNA library construction:
                Helix Research Institute (HRI); 5'-end one pass sequencing: HRI,
                Research Association for Biotechnology (RAB) and Biotechnology
                Center, National Institute of Technology and Evaluation; 3'-end one
                pass sequencing: RAB.
FEATURES   Location/Qualifiers
     source   1. .362
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="FCBBF3015237"
                     /tissue_type="brain"
                     /dev_stage="fetal"
                     /clone_lib="FCBBF3"
                     /note="Vector: pME18SFL3"
ORIGIN

  Query Match   16.1%; Score 96.4; DB 9; Length 362;
  Best Local Similarity 99.0%; Pred. No. 1.3e-06;
  Matches 97; Conservative   0; Mismatches   1; Indels   0; Gaps   0;

Qy   1 TACAATAGACCCTGCTTCTTTCAACGCTTCAACTGCTTTGCTATGTGTTACATCACGAAC 60
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   132 TACAATAGACCCTGCTTCTTTCAACGCTTCAACTGCTTTGCTATGTGTTACATCACGAAC 73

```
Qy       61 ATCTACTTCATTTACTTGTAATATACAGTCATTGACCC 98
            ||||||||||||||||| |||||||||||||||||||||
Db       72 ATCTACTTCATTTACTCGTAATATACAGTCATTGACCC 35
```

RESULT 6
AA831793
LOCUS   AA831793   482 bp   mRNA   linear   EST 05-MAR-1998
DEFINITION   oa59d10.s1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:1309267 3'
               similar to SW:SP97_HUMAN Q12959 PRESYNAPTIC PROTEIN SAP97 ;, mRNA
               sequence.
ACCESSION   AA831793
VERSION   AA831793.1 GI:2904892
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
  ORGANISM   Homo sapiens
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
               Hominidae; Homo.
REFERENCE   1 (bases 1 to 482)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE   National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
               Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT   Contact: Robert Strausberg, Ph.D.
               Email: cgapbs-r@mail.nih.gov
               Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,
               Ph.D., Gerald Marti, M.D.
                cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
               Bonaldo, Ph.D.
                cDNA Library Arrayed by: Greg Lennon, Ph.D.
                DNA Sequencing by: Washington University Genome Sequencing Center
                Clone distribution: NCI-CGAP clone distribution information can be
               found through the I.M.A.G.E. Consortium/LLNL at:
               www-bio.llnl.gov/bbrp/image/image.html
               Insert Length: 921   Std Error: 0.00
               Seq primer: -40m13 fwd. ET from Amersham
               High quality sequence stop: 467.
FEATURES   Location/Qualifiers
     source   1. .482
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1309267"
                     /tissue_type="germinal center B cell"
                     /lab_host="DH10B"
                     /clone_lib="NCI_CGAP_GCB1"
                     /note="Vector: pT7T3D-PacI; Site_1: Not I; Site_2: Eco RI;
                     1st strand cDNA was prepared from human tonsillar cells
                     enriched for germinal center B cells by flow sorting
                     (CD20+, IgD-), provided by Dr. Louis M. Staudt (NCI), Dr.
                     David Allman (NCI) and Dr. Gerald Marti (CBER). cDNA
                     synthesis was primed with a Not I - oligo(dT) primer
                     [5'-TGTTACCAATCTGAAGTGGGAGCGGCCGCCTCATTTTTTTTTTTTTTTTTT-3'
                     ]. Double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
```

and Eco RI sites of the modified pT7T3 vector. Library
                    went through one round of normalization, and was
                    constructed by Bento Soares and M. Fatima Bonaldo."
ORIGIN

    Query Match              16.1%;  Score 96.4;  DB 1;  Length 482;
    Best Local Similarity    99.0%;  Pred. No. 1.3e-06;
    Matches    97; Conservative     0; Mismatches     1; Indels     0; Gaps     0;


Qy        1 TACAATAGACCCTGCTTCTTTCAACGCTTCAACTGCTTTGCTATGTGTTACATCACGAAC 60
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       21 TACAATAGACCCTGCTTCTTTCAACGCTTCAACTGCTTTGCTATGTGTTACATCACGAAC 80


Qy       61 ATCTACTTCATTTACTTGTAATATACAGTCATTGACCC 98
            |||||||||||||||| ||||||||||||||||||||||
Db       81 ATCTACTTCATTTACTCGTAATATACAGTCATTGACCC 118



RESULT 7
BU741256
LOCUS         BU741256                   531 bp    mRNA     linear   EST 10-OCT-2002
DEFINITION    UI-E-EJ0-air-d-04-0-UI.s1 UI-E-EJ0 Homo sapiens cDNA clone
              UI-E-EJ0-air-d-04-0-UI 3', mRNA sequence.
ACCESSION     BU741256
VERSION       BU741256.1  GI:23684413
KEYWORDS      EST.
SOURCE        Homo sapiens (human)
  ORGANISM    Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
              Hominidae; Homo.
REFERENCE     1  (bases 1 to 531)
  AUTHORS     Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE       Normalization and subtraction: two approaches to facilitate gene
              discovery
  JOURNAL     Genome Res. 6 (9), 791-806 (1996)
   PUBMED     8889548
COMMENT       Contact: Soares, MB
              Coordinated Laboratory for Computational Genomics
              University of Iowa
              375 Newton Road , 4156  MEBRF, Iowa City, IA 52242, USA
              Tel: 319 335 8250
              Fax: 319 335 9565
              Email: bento-soares@uiowa.edu
              Tissue Procurement: Dr. Gregg Hageman
               cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
               cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
               DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
               Clone Distribution: Researchers may obtain clones from Research
              Genetics (www.resgen.com).
              Seq primer: M13 FORWARD
              POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1. .531
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"

```
                    /clone="UI-E-EJ0-air-d-04-0-UI"
                    /tissue_type="fetal eyes, lens, eye anterior segment,
                    optic nerve, retina, Retina Foveal and Macular, RPE and
                    Choroid"
                    /dev_stage="fetal and adult"
                    /lab_host="DH10B (Life Technologies) (T1 phage resistant)"
                    /clone_lib="UI-E-EJ0"
                    /note="Organ: eye; Vector: pT7T3-Pac (Pharmacia) with a
                    modified polylinker; Site_1: EcoR I; Site_2: Not I;
                    UI-E-EJ0 is a subtracted cDNA library constructed
                    according to Bonaldo, Lennon and Soares, Genome Research,
                    6:791-806, 1996. First strand cDNA synthesis was primed
                    with an oligo-dT primer containing a Not I site. Double
                    stranded cDNA was ligated to an EcoR I adaptor, digested
                    with Not I, and cloned directionally into pT7T3-Pac
                    vector. The oligonucleotide used to prime the synthesis of
                    first-strand cDNA contains a library tag sequence that is
                    located between the Not I site and the (dT)18 tail. The
                    sequence tags for this library are: fetal eyes,
                    AGAATCAAGA; lens, CGATTAGCGA; eye anterior segment,
                    AATGCCGCAT; optic nerve, CCATTAAGTG; retina, CCGCG; Retina
                    Foveal and Macular, GTCC; RPE and Choroid, ACCTA. This
                    library was created for the program, Gene Discovery in the
                    Visual System, supported by National Eye Institute (NEI).
                    TAG_TISSUE=human lens
                    TAG_LIB=UI-E-EJ0
                    TAG_SEQ=CGATTAGCGA"
ORIGIN


     Query Match              16.1%;   Score 96.4;   DB 3;   Length 531;
     Best Local Similarity    99.0%;   Pred. No. 1.3e-06;
       Matches    97; Conservative     0; Mismatches    1; Indels    0; Gaps    0;


Qy         1 TACAATAGACCCTGCTTCTTTCAACGCTTCAACTGCTTTGCTATGTGTTACATCACGAAC 60
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        30 TACAATAGACCCTGCTTCTTTCAACGCTTCAACTGCTTTGCTATGTGTTACATCACGAAC 89


Qy        61 ATCTACTTCATTTACTTGTAATATACAGTCATTGACCC 98
             |||||||||||||||| |||||||||||||||||||||
Db        90 ATCTACTTCATTTACTCGTAATATACAGTCATTGACCC 127



RESULT 8
BM727000/c
LOCUS           BM727000                   532 bp    mRNA      linear    EST 01-MAR-2002
DEFINITION      UI-E-EJ0-air-d-04-0-UI.r1 UI-E-EJ0 Homo sapiens cDNA clone
                UI-E-EJ0-air-d-04-0-UI 5', mRNA sequence.
ACCESSION       BM727000
VERSION         BM727000.1  GI:19048333
KEYWORDS        EST.
SOURCE          Homo sapiens (human)
  ORGANISM      Homo sapiens
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
                Hominidae; Homo.
REFERENCE       1  (bases 1 to 532)
  AUTHORS       Bonaldo,M.F., Lennon,G. and Soares,M.B.
```

```
TITLE        Normalization and subtraction: two approaches to facilitate gene
             discovery
JOURNAL      Genome Res. 6 (9), 791-806 (1996)
PUBMED       8889548
COMMENT      Contact: Soares, MB
             Coordinated Laboratory for Computational Genomics
             University of Iowa
             375 Newton Road , 4156  MEBRF, Iowa City, IA 52242, USA
             Tel: 319 335 8250
             Fax: 319 335 9565
             Email: bento-soares@uiowa.edu
             Tissue Procurement: Dr. Gregg Hageman
              cDNA Library preparation: Dr. M. Bento Soares, Univeristy of Iowa
              cDNA Library Arrayed by: Dr. M. Bento Soares, Univeristy of Iowa
              DNA Sequencing by: Dr. M. Bento Soares, Univeristy of Iowa
              Clone Distribution: Researchers may obtain clones from Research
             Genetics (www.resgen.com).
             Seq primer: M13 Reverse.
FEATURES             Location/Qualifiers
     source          1. .532
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="UI-E-EJ0-air-d-04-0-UI"
                     /tissue_type="fetal eyes, lens, eye anterior segment,
                     optic nerve, retina, Retina Foveal and Macular, RPE and
                     Choroid"
                     /dev_stage="fetal and adult"
                     /lab_host="DH10B (Life Technologies) (T1 phage resistant)"
                     /clone_lib="UI-E-EJ0"
                     /note="Organ: eye; Vector: pT7T3-Pac (Pharmacia) with a
                     modified polylinker; Site_1: EcoR I; Site_2: Not I;
                     UI-E-EJ0 is a subtracted cDNA library constructed
                     according to Bonaldo, Lennon and Soares, Genome Research,
                     6:791-806, 1996. First strand cDNA synthesis was primed
                     with an oligo-dT primer containing a Not I site. Double
                     stranded cDNA was ligated to an EcoR I adaptor, digested
                     with Not I, and cloned directionally into pT7T3-Pac
                     vector. The oligonucleotide used to prime the synthesis of
                     first-strand cDNA contains a library tag sequence that is
                     located between the Not I site and the (dT)18 tail. The
                     sequence tags for this library are: fetal eyes,
                     AGAATCAAGA; lens, CGATTAGCGA; eye anterior segment,
                     AATGCCGCAT; optic nerve, CCATTAAGTG; retina, CCGCG; Retina
                     Foveal and Macular, GTCC; RPE and Choroid, ACCTA. This
                     library was created for the program, Gene Discovery in the
                     Visual System, supported by National Eye Institute (NEI)."
ORIGIN

    Query Match              16.1%;  Score 96.4;  DB 3;  Length 532;
    Best Local Similarity    99.0%;  Pred. No. 1.3e-06;
    Matches    97; Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        1 TACAATAGACCCTGCTTCTTTCAACGCTTCAACTGCTTTGCTATGTGTTACATCACGAAC 60
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      493 TACAATAGACCCTGCTTCTTTCAACGCTTCAACTGCTTTGCTATGTGTTACATCACGAAC 434
```

```
Qy        61 ATCTACTTCATTTACTTGTAATATACAGTCATTGACCC 98
             ||||||||||||||||| |||||||||||||||||||||
Db       433 ATCTACTTCATTTACTCGTAATATACAGTCATTGACCC 396
```

RESULT 9
DA315109/c
LOCUS   DA315109   578 bp   mRNA   linear   EST 30-OCT-2005
DEFINITION   DA315109 BRHIP3 Homo sapiens cDNA clone BRHIP3005378 5', mRNA
              sequence.
ACCESSION   DA315109
VERSION   DA315109.1 GI:78306086
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
  ORGANISM   Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
              Hominidae; Homo.
REFERENCE   1 (bases 1 to 578)
  AUTHORS   Kimura,K., Wakamatsu,A., Suzuki,Y., Ota,T., Nishikawa,T.,
              Yamashita,R., Yamamoto,J., Sekine,M., Tsuritani,K., Wakaguri,H.,
              Ishii,S., Sugiyama,T., Saito,K., Isono,Y., Irie,R., Kushida,N.,
              Yoneyama,T., Otsuka,R., Kanda,K., Yokoi,T., Kondo,H., Wagatsuma,M.,
              Murakawa,K., Ishida,S., Ishibashi,T., Takahashi-Fujii,A.,
              Tanase,T., Nagai,K., Kikuchi,H., Nakai,K., Isogai,T. and Sugano,S.
  TITLE   Diversification of Transcriptional Modulation: Large-scale
              Identification and Characterization of Putative Alternative
              Promoters of Human Genes
  JOURNAL   Genome Res. 16 (1), 55-65 (2006)
  PUBMED   16344560
COMMENT   Contact: Takao Isogai
              FLJ Project (HRI Team)
              Helix Research Institute
              2-6-7 Kazusa-Kamatari, Kisarazu, Chiba, 292-0818, Japan
              Tel: 81-438-52-3975
              Fax: 81-438-52-3986
              Email: flj-cdna@nifty.com
              NEDO human cDNA project (New Energy and Industrial Technology
              Developmental Organization, Japan); cDNA library construction:
              Helix Research Institute (HRI); 5'-end one pass sequencing: HRI,
              Research Association for Biotechnology (RAB) and Biotechnology
              Center, National Institute of Technology and Evaluation; 3'-end one
              pass sequencing: RAB.
FEATURES   Location/Qualifiers
     source   1. .578
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="BRHIP3005378"
                     /tissue_type="hippocampus"
                     /clone_lib="BRHIP3"
                     /note="Vector: pME18SFL3"
ORIGIN

   Query Match   16.1%; Score 96.4; DB 9; Length 578;
   Best Local Similarity 99.0%; Pred. No. 1.3e-06;
   Matches   97; Conservative   0; Mismatches   1; Indels   0; Gaps   0;
```

```
Qy        1 TACAATAGACCCTGCTTCTTTCAACGCTTCAACTGCTTTGCTATGTGTTACATCACGAAC 60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      264 TACAATAGACCCTGCTTCTTTCAACGCTTCAACTGCTTTGCTATGTGTTACATCACGAAC 205

Qy       61 ATCTACTTCATTTACTTGTAATATACAGTCATTGACCC 98
            ||||||||||||||| |||||||||||||||||||||
Db      204 ATCTACTTCATTTACTCGTAATATACAGTCATTGACCC 167
```

RESULT 10
DA262508/c

```
LOCUS       DA262508                 580 bp    mRNA    linear   EST 03-NOV-2005
DEFINITION  DA262508 BRCAN2 Homo sapiens cDNA clone BRCAN2015005 5', mRNA
            sequence.
ACCESSION   DA262508
VERSION     DA262508.1  GI:79168489
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 580)
  AUTHORS   Kimura,K., Wakamatsu,A., Suzuki,Y., Ota,T., Nishikawa,T.,
            Yamashita,R., Yamamoto,J., Sekine,M., Tsuritani,K., Wakaguri,H.,
            Ishii,S., Sugiyama,T., Saito,K., Isono,Y., Irie,R., Kushida,N.,
            Yoneyama,T., Otsuka,R., Kanda,K., Yokoi,T., Kondo,H., Wagatsuma,M.,
            Murakawa,K., Ishida,S., Ishibashi,T., Takahashi-Fujii,A.,
            Tanase,T., Nagai,K., Kikuchi,H., Nakai,K., Isogai,T. and Sugano,S.
  TITLE     Diversification of Transcriptional Modulation: Large-scale
            Identification and Characterization of Putative Alternative
            Promoters of Human Genes
  JOURNAL   Genome Res. 16 (1), 55-65 (2006)
   PUBMED   16344560
COMMENT     Contact: Takao Isogai
            FLJ Project (HRI Team)
            Helix Research Institute
            2-6-7 Kazusa-Kamatari, Kisarazu, Chiba, 292-0818, Japan
            Tel: 81-438-52-3975
            Fax: 81-438-52-3986
            Email: flj-cdna@nifty.com
            NEDO human cDNA project (New Energy and Industrial Technology
            Developmental Organization, Japan); cDNA library construction:
            Helix Research Institute (HRI); 5'-end one pass sequencing: HRI,
            Research Association for Biotechnology (RAB) and Biotechnology
            Center, National Institute of Technology and Evaluation; 3'-end one
            pass sequencing: RAB.
FEATURES             Location/Qualifiers
     source          1. .580
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="BRCAN2015005"
                     /tissue_type="caudate nucleus"
                     /clone_lib="BRCAN2"
                     /note="Vector: pME18SFL3"
```

ORIGIN

```
     Query Match              16.1%;   Score 96.4;   DB 9;   Length 580;
     Best Local Similarity    99.0%;   Pred. No. 1.3e-06;
     Matches   97;  Conservative   0;  Mismatches   1;  Indels   0;  Gaps   0;


Qy        1 TACAATAGACCCTGCTTCTTTCAACGCTTCAACTGCTTTGCTATGTGTTACATCACGAAC 60
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      218 TACAATAGACCCTGCTTCTTTCAACGCTTCAACTGCTTTGCTATGTGTTACATCACGAAC 159


Qy       61 ATCTACTTCATTTACTTGTAATATACAGTCATTGACCC 98
            ||||||||||||||||| ||||||||||||||||||||
Db      158 ATCTACTTCATTTACTCGTAATATACAGTCATTGACCC 121
```

RESULT 11
BF677732/c
```
LOCUS       BF677732                842 bp    mRNA    linear   EST 21-DEC-2000
DEFINITION  602085430F1 NIH_MGC_83 Homo sapiens cDNA clone IMAGE:4249854 5',
            mRNA sequence.
ACCESSION   BF677732
VERSION     BF677732.1  GI:11951627
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 842)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: CLONETECH Laboratories, Inc.
             cDNA Library Preparation: CLONETECH Laboratories, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1070  row: k  column: 07
            High quality sequence stop: 614.
FEATURES             Location/Qualifiers
     source          1..842
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4249854"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /clone_lib="NIH_MGC_83"
                     /note="Organ: prostate; Vector: pDNR-LIB (Clontech);
                     Site_1: SfiI (ggccgcctcggcc); Site_2: SfiI
                     (ggccattatggcc); 5' and 3' adaptors were used in cloning
                     as follows: 5' adaptor sequence:  5'-CACGGCCATTATGGCC-3'
                     and 3' adaptor sequence:
                     5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3' (where B = A,
```

C, or G and N = A, C, G, or T). Average insert size 1.4
kb (range 0.5-4.0 kb). 14/15 colonies contained inserts
by PCR. This library was enriched for full-length clones
and was constructed by Clontech Laboratories (Palo Alto,
CA)."
ORIGIN

```
   Query Match            16.1%;   Score 96.4;   DB 7;   Length 842;
   Best Local Similarity  99.0%;   Pred. No. 1.2e-06;
   Matches  97;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;


Qy      1 TACAATAGACCCTGCTTCTTTCAACGCTTCAACTGCTTTGCTATGTGTTACATCACGAAC 60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    303 TACAATAGACCCTGCTTCTTTCAACGCTTCAACTGCTTTGCTATGTGTTACATCACGAAC 244


Qy     61 ATCTACTTCATTTACTTGTAATATACAGTCATTGACCC 98
          ||||||||||||||||| ||||||||||||||||||||
Db    243 ATCTACTTCATTTACTCGTAATATACAGTCATTGACCC 206
```

RESULT 12
BG164149/c
LOCUS   BG164149   860 bp   mRNA   linear   EST 06-FEB-2001
DEFINITION   602343020F1 NIH_MGC_89 Homo sapiens cDNA clone IMAGE:4453276 5',
             mRNA sequence.
ACCESSION   BG164149
VERSION   BG164149.1 GI:12670852
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
             Hominidae; Homo.
REFERENCE   1 (bases 1 to 860)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
             Email: cgapbs-r@mail.nih.gov
             Tissue Procurement: ATCC
              cDNA Library Preparation: Life Technologies, Inc.
              cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
              Clone distribution: MGC clone distribution information can be
             found through the I.M.A.G.E. Consortium/LLNL at:
             http://image.llnl.gov
             Plate: LLAM10243 row: g column: 05
             High quality sequence stop: 687.
FEATURES   Location/Qualifiers
     source   1..860
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4453276"
                     /tissue_type="hypernephroma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_89"

```
                    /note="Organ: kidney; Vector: pCMV-SPORT6; Site_1: NotI;
                    Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                    Average insert size 1.3 kb. Library enriched for
                    full-length clones and constructed by Life Technologies.
                    Note: this is a NIH_MGC Library."
ORIGIN

     Query Match             16.1%;  Score 96.4;  DB 2;  Length 860;
     Best Local Similarity   99.0%;  Pred. No. 1.2e-06;
       Matches  97;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 TACAATAGACCCTGCTTCTTTCAACGCTTCAACTGCTTTGCTATGTGTTACATCACGAAC 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        232 TACAATAGACCCTGCTTCTTTCAACGCTTCAACTGCTTTGCTATGTGTTACATCACGAAC 173

Qy         61 ATCTACTTCATTTACTTGTAATATACAGTCATTGACCC 98
              ||||||||||||||||| ||||||||||||||||||||
Db        172 ATCTACTTCATTTACTCGTAATATACAGTCATTGACCC 135


RESULT 13
AY420526/c
LOCUS       AY420526                2675 bp    DNA        linear   GSS 17-DEC-2003
DEFINITION  Pan troglodytes DLG1 gene, VIRTUAL TRANSCRIPT, partial sequence,
            genomic survey sequence.
ACCESSION   AY420526
VERSION     AY420526.1  GI:39776483
KEYWORDS    GSS.
SOURCE      Pan troglodytes (chimpanzee)
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Pan.
REFERENCE   1  (bases 1 to 2675)
  AUTHORS   Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
            Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
            Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
            Adams,M.D. and Cargill,M.
  TITLE     Inferring nonneutral evolution from human-chimp-mouse orthologous
            gene trios
  JOURNAL   Science 302 (5652), 1960-1963 (2003)
  PUBMED    14671302
REFERENCE   2  (bases 1 to 2675)
  AUTHORS   Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
            Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
            Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
            Adams,M.D. and Cargill,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
            Rockville, MD 20850, USA
COMMENT     This sequence as made by sequencing genomic exons and ordering them
            based on alignment.
FEATURES             Location/Qualifiers
     source          1. .2675
                     /organism="Pan troglodytes"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9598"
```

```
           gene              <1. .>2675
                             /gene="DLG1"
                             /locus_tag="HCM7253"
       ORIGIN

         Query Match            16.1%;   Score 96.4;  DB 14;  Length 2675;
         Best Local Similarity  99.0%;   Pred. No. 1.1e-06;
         Matches   97;  Conservative     0;  Mismatches    1;  Indels    0;  Gaps    0;

       Qy        1 TACAATAGACCCTGCTTCTTTCAACGCTTCAACTGCTTTGCTATGTGTTACATCACGAAC 60
                    |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
       Db      890 TACAATAGACCCTGCTTCTTTCAACGCTTCAACTGCTTTGCTATGTGTTACATCACGAAC 831

       Qy       61 ATCTACTTCATTTACTTGTAATATACAGTCATTGACCC 98
                    |||||||||||||||| ||||||||||||||||||||
       Db      830 ATCTACTTCATTTACTCGTAATATACAGTCATTGACCC 793


       RESULT 14
       AA480937
       LOCUS       AA480937                447 bp    mRNA    linear    EST 14-AUG-1997
       DEFINITION  aa28c06.r1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:814570 5'
                   similar to TR:G558438 G558438 HOMOLOG OF DROSOPHILA DISCS LARGE
                   PROTEIN, ISOFORM 1. ;, mRNA sequence.
       ACCESSION   AA480937
       VERSION     AA480937.1  GI:2210489
       KEYWORDS    EST.
       SOURCE      Homo sapiens (human)
         ORGANISM  Homo sapiens
                   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                   Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
                   Hominidae; Homo.
       REFERENCE   1  (bases 1 to 447)
         AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
         TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
                   Tumor Gene Index
         JOURNAL   Unpublished (1997)
       COMMENT     Contact: Robert Strausberg, Ph.D.
                   Email: cgapbs-r@mail.nih.gov
                   Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,
                   Ph.D., Gerald Marti, M.D.
                    cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
                   Bonaldo, Ph.D.
                    cDNA Library Arrayed by: Greg Lennon, Ph.D.
                    DNA Sequencing by: Washington University Genome Sequencing Center
                    Clone distribution: NCI-CGAP clone distribution information can be
                   found through the I.M.A.G.E. Consortium/LLNL at:
                   www-bio.llnl.gov/bbrp/image/image.html

                   Possible reversed clone: similarity on wrong strand
                   Seq primer: -28m13 rev1 ET from Amersham
                   High quality sequence stop: 385.
       FEATURES            Location/Qualifiers
            source         1. .447
                           /organism="Homo sapiens"
                           /mol_type="mRNA"
                           /db_xref="GDB:6032940"
```

```
                                /db_xref="taxon:9606"
                                /clone="IMAGE:814570"
                                /tissue_type="germinal center B cell"
                                /lab_host="DH10B"
                                /clone_lib="NCI_CGAP_GCB1"
                                /note="Vector: pT7T3D-PacI; Site_1: Not I; Site_2: Eco RI;
                                1st strand cDNA was prepared from human tonsillar cells
                                enriched for germinal center B cells by flow sorting
                                (CD20+, IgD-), provided by Dr. Louis M. Staudt (NCI), Dr.
                                David Allman (NCI) and Dr. Gerald Marti (CBER). cDNA
                                synthesis was primed with a Not I - oligo(dT) primer
                                [5'-TGTTACCAATCTGAAGTGGGAGCGGCCGCCTCATTTTTTTTTTTTTTTTT-3'
                                ]. Double-stranded cDNA was ligated to Eco RI adaptors
                                (Pharmacia), digested with Not I and cloned into the Not I
                                and Eco RI sites of the modified pT7T3 vector. Library
                                went through one round of normalization, and was
                                constructed by Bento Soares and M. Fatima Bonaldo."
ORIGIN

     Query Match              15.8%;  Score 94.8;  DB 1;  Length 447;
     Best Local Similarity    98.0%;  Pred. No. 2.5e-06;
        Matches   96;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 TACAATAGACCCTGCTTCTTTCAACGCTTCAACTGCTTTGCTATGTGTTACATCACGAAC 60
              |||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 TACAGTAGACCCTGCTTCTTTCAACGCTTCAACTGCTTTGCTATGTGTTACATCACGAAC 100

Qy         61 ATCTACTTCATTTACTTGTAATATACAGTCATTGACCC 98
              |||||||||||||||| |||||||||||||||||||||
Db        101 ATCTACTTCATTTACTCGTAATATACAGTCATTGACCC 138


RESULT 15
BP229026/c
LOCUS           BP229026                 582 bp    mRNA    linear    EST 15-SEP-2004
DEFINITION      BP229026 Sugano cDNA library, fetal brain Homo sapiens cDNA clone
                FBR01745, mRNA sequence.
ACCESSION       BP229026
VERSION         BP229026.1  GI:52101936
KEYWORDS        EST.
SOURCE          Homo sapiens (human)
  ORGANISM      Homo sapiens
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
                Hominidae; Homo.
REFERENCE       1  (bases 1 to 582)
  AUTHORS       Suzuki,Y., Yamashita,R., Shirota,M., Sakakibara,Y., Chiba,J.,
                Mizushima-Sugano,J., Nakai,K. and Sugano,S.
  TITLE         Sequence comparison of human and mouse genes reveals a homologous
                block structure in the promoter regions
  JOURNAL       Genome Res. 14 (9), 1711-1718 (2004)
   PUBMED       15342556
COMMENT         Contact: Yutaka Suzuki
                Department of Virology
                Institute of Medical Science, University of Tokyo
                4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
                Tel: 81-3-5449-5343
```